

Query Match 100.0%; Score 20; DB 23; Length 346;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagcagcagagac 20
|
Db 140 gtggccgagcagcagagac 159

RESULT 4

AAS91261
ID AAS91261 standard; cDNA: 346 BP.

XX AAS91261;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #27065.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG27074.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID NO 27065; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 346 BP; 46 A; 116 C; 133 G; 51 T; 0 other;

Query Match 100.0%; Score 20; DB 23; Length 346;

Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagcagcagagac 20
|
Db 140 gtggccgagcagcagagac 159

RESULT 5

AAS87785
ID AAS87785 standard; cDNA: 370 BP.

XX AAS87785;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #23589.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG23598.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID NO 23589; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 370 BP; 58 A; 119 C; 137 G; 56 T; 0 other;

Query Match 100.0%; Score 20; DB 23; Length 370;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcccgcgcgcgcgcgcgcgc 20
|||||
Db 164 gtgcccgcgcgcgcgcgcgcgc 183

RESULT 6

AAS91972
ID AAS91972 standard; cDNA: 370 BP.

XX
AC AAS91972;

XX
DT 13-FEB-2002 (first entry)

XX
DE DNA encoding novel human diagnostic protein #27776.

XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX
OS Homo sapiens.

XX
PN WO200175067-A2.

XX
PD 11-OCT-2001.

XX
PE 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.

XX
PR 23-AUG-2000; 2000US-0649167.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Drmanac RT, Liu C, Tang YT;

XX
DR WPI: 2001-639362/73.

XX
DR P-PSDB: ABG27785.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1: SEQ ID No 27776; 103bp; English.

XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX
SQ Sequence 370 BP; 58 A; 119 C; 137 G; 56 T; 0 other;

Query Match 100.0%; Score 20; DB 23; Length 370;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcccgcgcgcgcgcgcgcgc 20
|||||
Db 164 gtgcccgcgcgcgcgcgcgcgc 183

RESULT 7

AAV06551/c
ID AAV06551 standard; DNA: 516 BP.

XX
AC AAV06551;

XX
DT 06-JUL-1998 (first entry)

XX
DE SCA2 gene fragment including CAG repeat region.

XX
KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;

XX
KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.

XX
OS Homo sapiens.

XX
FH Key

FT complement (241..257)

FT /tag= a

FT /note= "primer SCA2-A binding site"

FT /tag= b

FT /note= "primer SCA2-B binding site"

FT /tag= c

FT /note= "predicted splice site"

FT /tag= d

FT /note= "CAG repeat region"

FT /tag= e

FT /note= "CAG repeat"

FT /tag= f

FT /note= "CAG repeat"

FT /tag= g

FT /note= "CAG repeat"

FT /tag= h

FT /note= "CAG repeat"

FT /tag= i

FT /note= "CAG repeat"

FT /tag= j

FT /note= "CAG repeat"

FT /tag= k

FT /note= "CAG repeat"

FT /tag= l

FT /note= "CAG repeat"

FT /tag= m

FT /note= "CAG repeat"

FT /tag= n

FT /note= "CAG repeat"

FT /tag= o

FT /note= "CAG repeat"

FT /tag= p

FT /note= "CAG repeat"

FT /tag= q

FT /note= "CAG repeat"

FT /tag= r

FT /note= "CAG repeat"

FT /tag= s

FT /note= "CAG repeat"

FT /tag= t

FT /note= "CAG repeat"

FT /tag= u

FT /note= "CAG repeat"

FT /tag= v

FT /note= "CAG repeat"

FT /tag= w

FT /note= "CAG repeat"

FT /tag= x

FT /note= "CAG repeat"

FT /tag= y

FT /note= "CAG repeat"

FT /tag= z

FT /note= "CAG repeat"

Query Match	100.0%	Score 20;	DB 19;	Length 516;
Best Local Similarity	100.0%	Pred. No. 19;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

0y	1	gtggccgagagcagagac	20
db	442	gtggccgagagcagagac	423

```

RESULT      8
AAV17229/c
ID          AAV17229 standard; DNA; 623 BP.
XX
AC          AAV17229;
XX
DT          29-JUN-1998 (first entry)
XX
DE          SCA2 gene fragment.
XX
KW          SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX
OS          Synthetic.
XX
FT          Key                      Location/Qualifiers
FT          CDS                      341..583
FT                                     /*tag=
FT                                     a
FT                                     /note= "SCA2 protein fragment, no stop codon given"
XX
XX          MO9803679-A1.
XX
XX          29-JAN-1998.
XX
XX          18-JUL-1996; 96WO-JP01999.
XX
XX          18-JUL-1996; 96WO-JP01999.
XX
XX          (SRLS-) SRL INC.
XX
XX          Sanpei K, Tsuji S;
XX
XX          WPI; 1998-120796/11.
XX          DR          P-PSDB; AAW41372.
XX
XX          Diagnosing spinocerebellar ataxis type II - by PCR and determining
XX          number of CAG repeat units
XX
XX          Example 1; Page 11-12; 23pp; Japanese.
XX
XX          This sequence represents a fragment of the SCA2 gene. It can be used in
XX          the method of the invention for diagnosing spinocerebellar ataxis type
XX          II, by performing PCR on the test DNA using two primers hybridising to
XX          parts of the SCA2 gene sequence, and determining the number of CAG
XX          repeats in the amplified products. The method provides an easy means for
XX          the diagnosis of spinocerebellar ataxis type II.
XX
XX          Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;
XX
XX
XX          Query Match                      100.0%; Score 20; DB 19; Length 623;
XX          Best Local Similarity 100.0%; Pred. No. 19;
XX          Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX          1 gfgccgaagacagagagac 20
XX          ||||||||||||||||
XX          528 GTGGCCGAGACAGACAGAC 509
XX
XX
XX          RESULT      9
XX          ID          AAT78912/c
XX          AAT78912 standard; cDNA; 4200 BP.
XX
XX          AAT78912;
XX
XX          09-FEB-1998 (first entry)
XX
XX          Spinocerebellar ataxia gene SCA2.
XX
XX          Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;
XX          repeat region; affinity; RNA binding protein; Kennedy disease;
XX          transcription initiation factor; lymphoblastic cell line; schizophrenia;

```

KM	Huntington's disease; dominant autosomal spinocerebellar ataxia;
KM	x-linked spino-bulbar muscular atrophy; familial spastic paraplegia;
KM	dentatorubral-pallidoluysian atrophy; bipolar affective disorder;
KM	manic depressive psychosis; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	3..2747
FT	/*tag- a
FT	/product= SCA2 protein
FT	/note= "this CDS contains a putative translational start
FT	codon for the SCA2 protein at positions 243-245
FT	2594..3640
FT	/*tag- b
FT	/note= "this second open reading frame may be derived
FT	by a frameshift or by alternative splicing"
FT	3..242
FT	/*tag- c
FT	/note= "putative open reading frame which is in frame
FT	with the putative translational start site of
FT	the SCA2 open reading frame"
FT	misc_signal
FT	239..245
FT	/*tag- d
FT	/note= "putative Kozak consensus signal"
FT	258..323
FT	/*tag= e
FT	/note= "encodes polyglutamine repeat region; contains
FT	repeats of CAG with 2 CAA codons interspersed"
FT	258..260
FT	/*tag- f
FT	/note= "CAG repeats"
FT	1..3986
FT	/*tag= g
FT	/note= "sequence contained in DANI clone"
FT	3987..4200
FT	/*tag- h
FT	/note= "derived from the EST's AAH92640, AAN90240 and
FT	AAZ13574 from dbEST database"
FT	4023..4029
FT	/*tag- i
FT	/note= "region which differs in length between the
FT	sequences of the EST clones AAH92640, AAN90240
FT	and AAZ13574"
XX	
PN	W09717445-A1.
XX	
PD	15-MAY-1997.
XX	
PE	08-NOV-1996; 96MO-FR01773.
XX	
PR	10-NOV-1995; 95FR-0013576.
XX	
PA	(CNRS) CNRS CENT NAT RECH SCI.
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI	Lutz Y, Mandel J, Tora L, Trottier Y;
XX	
DR	WPI: 1997-281034/25.
DR	P-PSDB: AAM24800, AAM24801.
XX	
PT	Antibody IC2 used for treating or preventing neuro-degenerative
PT	diseases - associated with proteins containing long poly:glutamine
PT	repeats, e.g. Huntington's disease
XX	
PS	Claim 21: Page 45-47; 69pp: French.
XX	
CC	The invention relates to a monoclonal antibody (MAB) IC2 for the
CC	treatment of neurodegenerative diseases associated with the presence
CC	of polyglutamine repeat regions. This MAB is already known for its
CC	affinity to the TATA binding protein (TBP) transcription initiation
CC	factor, especially at the amino acid sequence LEEQGRQDQDQD found at
CC	the N-terminus of TBP. MAB IC2 has been shown to have a high affinity

CC		for polyglutamine repeats with a proportional affinity to the number
CC		of glutamine repeats.. This affinity has been used to identify genes
CC		encoding proteins containing long polyglutamine repeats which are
CC		implicated in neurodegenerative diseases. A screen of an expression
CC		library, generated from a lymphoblastic cell line from a patient
CC		suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
CC		new sequences (AA78906-T78911) encoding polyglutamine repeats. Mab 1C2
CC		also isolated the complete SCA2 gene in clone DAN1 (sequence presented
CC		here). The sequence appears to contain 2 open reading frames (ORF) the
CC		second of which may be generated by a frameshift slippage or by an
CC		alternative splicing event. The first ORF also encodes a 22 amino acid
CC		polyglutamine repeat region near the N-terminus of the protein. Normal
CC		SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
CC		interspersed whereas the mutant sequence from patients with SCA
CC		contains at least 30, preferably 37-50 CAG repeats.
CC		Mab 1C2, active fragment of it or nucleic acids encoding it are
CC		specifically used to treat Huntington's disease, SCA types 1-5 or 7,
CC		X-linked spinocerebellar muscular atrophy (Kennedy disease),
CC		dentatorubral-pallidoluysial atrophy, dominant autosomal spinocerebellar
CC		ataxia, familial spastic paraplegia, bipolar affective disorder, manic
CC		depressive psychoses and schizophrenia.
SQ		Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;
OY		Query Match
	Best Local Similarity	100.0%; Score 20; DB 18; Length 4200;
	Matches 20; Conservative	100.0%; Pred. NO. 18; Mismatches 0; Indels 0; Gaps 0;
DB		1 GTGCGGAGGACGAGGAC 20 433 GTGCCGAGGACGAGGAC 414
RESULT 10		
ID	AAV30270/C	
XX	AAV30270 standard; DNA; 4367 BP.	
AC	AAV30270;	
XX		
DT	02-OCT-1998 (first entry)	
XX		
DE	Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.	
KX	Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;	
KW	CAG repeat; neurodegenerative disease; ds.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	CDS	49..3990
FT	/tag= a	
FT	/product= "Spinocerebellar ataxia type 2 associated	
FT	repeat_region	544..612
FT	/tag= b	
FT	/note= "normal CAG repeat region; this is increased in	
FT	repeat_unit	544..546
FT	/tag= c	
XX		
PN	W09818920-A1.	
XX		
PD	07-MAY-1998.	
XX		
PF	30-OCT-1997;	97MO-JP03946.
XX		
PR	30-OCT-1996;	96JP-0304059.
XX		
PA	(SRLS-) SRL INC.	
PI	Sanpel K, Tsujl S;	
XX		

DR WPI: 1998-272215/24.
 DR P-PSDB: AAM60213.
 XX Nucleic acid fragments associated with spinocerebellar ataxia type 2
 PT - contain increased number of CAG repeat region compared to normal
 PT gene
 XX
 PS Claim 1: Pages 13-22; 38pp; Japanese.
 XX
 CC This represents the sequence of a gene causative of spinocerebellar
 CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
 CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
 CC expression product produces a polyglutamine sequence from Gln-166 to
 CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
 CC patients this number is increased to 35-100. Peptides encoded by nucleic
 CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
 CC gene, antibodies recognising the peptides and antisense nucleic acids
 CC hybridising with the nucleic acid fragments can be used for the
 CC investigation and diagnosis of SCA2. They can also be used for the
 CC treatment of SCA2 by antisense therapy or gene therapy.
 CC
 SQ Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;
 Query Match 100.0%; Score 20; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 gtagccgagagagagagac 20
 DB 722 GTGCCGACGAGAGAGAC 703
 RESUT 11
 ID AAV06552/c
 XX AAV06552 standard; cDNA: 4481 BP.
 AC AAV06552;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human SCA2 cDNA including CAG repeat region.
 XX
 KM SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KM diagnosis; olivoponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 XX
 FH Location/Qualifiers
 FT CDS
 FT 164..4101
 FT /tag= a
 FT complement (631..648)
 FT /tag= b
 FT /note= "primer SCA2-A binding site"
 FT 740..757
 FT /tag= c
 FT /note= "primer SCA2-B binding site"
 FT 1070..1091
 FT /tag= d
 FT /note= "primer SCA2-14B binding site"
 FT 899..900
 FT /tag= e
 FT /note= "predicted splice site"
 FT 658..723
 FT /tag= f
 FT /note= "CAG repeat region"
 FT 658..660
 FT /tag= g
 FT /note= "CAG repeat"
 FT 661..663
 FT /tag= h
 FT /note= "CAG repeat"
 FT 664..666
 FT repeat_unit

FT /tag= i
 FT /note= "CAG repeat"
 FT 667..669
 FT /tag= j
 FT /note= "CAG repeat"
 FT 670..672
 FT /tag= k
 FT /note= "CAG repeat"
 FT 673..675
 FT /tag= l
 FT /note= "CAG repeat"
 FT 676..678
 FT /tag= m
 FT /note= "CAG repeat"
 FT 679..681
 FT /tag= n
 FT /note= "CAG repeat"
 FT 685..687
 FT /tag= o
 FT /note= "CAG repeat"
 FT 688..690
 FT /tag= p
 FT /note= "CAG repeat"
 FT 691..693
 FT /tag= q
 FT /note= "CAG repeat"
 FT 694..696
 FT /tag= r
 FT /note= "CAG repeat"
 FT 700..702
 FT /tag= s
 FT /note= "CAG repeat"
 FT 703..705
 FT /tag= t
 FT /note= "CAG repeat"
 FT 706..708
 FT /tag= u
 FT /note= "CAG repeat"
 FT 709..711
 FT /tag= v
 FT /note= "CAG repeat"
 FT 712..714
 FT /tag= w
 FT /note= "CAG repeat"
 FT 715..717
 FT /tag= x
 FT /note= "CAG repeat"
 FT 718..720
 FT /tag= y
 FT /note= "CAG repeat"
 FT 721..723
 FT /tag= z
 FT /note= "CAG repeat"
 FT
 PN WO9742314-A1.
 PD
 XX
 PD 13-NOV-1997.
 XX
 PE 08-MAY-1997; 97WO-US07725.
 XX
 PR 08-OCT-1996; 96US-0727084.
 PR 08-MAY-1996; 96US-0017388.
 PR 19-JUL-1996; 96US-0022207.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst S;
 XX
 DR WPI: 1998-086523/08.
 DR P-PSDB: AAM33807.
 XX
 PT Nucleic acids encoding human and mouse ataxin 2 - a product of the
 PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of

PT ataxia type 2
XX
PS Claim 6; Page 52-58; 98pp; English.
XX
CC This cDNA sequence corresponds to a novel SCA2 gene encoding a human
CC spinocerebellar ataxis-2 (SCA2) polypeptide, designated ataxin-2
CC (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
CC human frontal cortex cDNA library in lambda ZapII were screened
CC with probes obtained by PCR amplification of plasmid AAP65122B (see
CC AAW06551). PCR products were used to screen the human adult frontal
CC cortex library, and 5' clones were obtained by RT-PCR of placental
CC mRNAs. Overlapping clones was used to generate the composite 4481
CC bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
CC or transcribed mRNA sequence in an individual having an expanded
CC CAG repeat at a location corresponding to the CAG repeat region of
CC the SCA2 gene. The presence of at least 13 CAG repeats above the
CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
CC Primers (see AAW9640-41) amplifying at least this region are used
CC for diagnosis. Also claimed are kits for detecting mutations at
CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
CC useful for studying the physiological roles of ataxin-2 and its
CC effect upon behaviour.
CC
XX
SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 20; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagagcagagac 20
|||||
DB 833 GTGGCCGAGGACGAGGAC 814

RESULT 12
AAZ23428/C
ID AAZ23428 standard; DNA: 4481 BP.
XX
AC AAZ23428;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human SCA2 DNA.
XX
KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidolysian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic; ss.
XX
OS Homo sapiens.
XX
PH key Location/Qualifiers
FT 163..4101
FT CDS /*tag= a
FT /product= "SCA2"
XX
PN WO945944-A1.
XX
PD 16-SEP-1999.
XX
PF 11-MAR-1999; 99WO-US05250.
XX
PR 12-MAR-1998; 98US-0041886.
XX
PA (BURN-) BURHAM INST.
XX
PI Bredesen DE, Rabizadeh S;
XX
DR WPI: 1999-561617/47.
DR P-PSDB; AAY33495.

XX
PT New proapoptotic dependence peptides, used to develop products for
PT treating, e.g. Alzheimer's disease -
PS
XX
PS Disclosure: Page 130-135; 199pp; English.
XX
CC This invention describes novel pure proapoptotic dependence peptides
CC which comprise a sequence of an active dependence domain selected from
CC dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC of inducing cell death and can be used to develop products to mediate or
CC inhibit apoptosis. The methods can be used for reducing the severity of
CC a proapoptotic dependence domain mediated pathological conditions e.g.
CC Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC spinocerebellar ataxias, dentatorubropallidolysian atrophy,
CC Machado-Joseph disease, stroke or head trauma. They can also be used for
CC reducing the severity of a pathological condition mediated by upregulated
CC cell proliferation or cell survival e.g. neoplastic, malignant,
CC autoimmune or fibrotic conditions. This sequence encodes the human
CC SCA2 polypeptide described in the method of the invention.
CC
XX
SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 4481;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagagcagagac 20
|||||
DB 833 GTGGCCGAGGACGAGGAC 814

RESULT 13
AAA28844
ID AAA28844 standard; DNA: 46 BP.
XX
AC AAA28844;
XX
DT 29-AUG-2000 (first entry)
XX
DE 3' primer for mature extracellular domain of FGFR5 DNA.
XX
KW FGFR-5; fibroblast growth factor receptor 5; cytosstatic; anti-sclerotic;
KW immunomodulatory; gastrointestinal; virulence; anti-inflammatory;
KW anti-ischemic; anti-atherosclerosis; angiogenic; endocrine;
KW anti-diabetic; gene therapy; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200024756-A1.
XX
PD 04-MAY-2000.
XX
PF 17-JUN-1999; 99WO-US13620.
XX
PR 23-OCT-1998; 98US-0105465.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE;
XX
DR WPI: 2000-387035/33.
XX
PT Nucleic acids encoding fibroblast growth factor-5 useful for the
PT prevention, diagnosis and treatment of conditions associated with
PT tissue repair and aberrant cell functions, e.g. cell survival and
PT proliferation
XX
PS Example 5; Page 101; 182pp; English.
XX
CC AAA28843-44 are primers used to amplify the mature extracellular domain

CC of fibroblast growth factor receptor protein, FGFR-5 DNA in a bacterial
CC expression vector. The FGFR-5 protein and DNA may be used in the
CC prevention, treatment and diagnosis of diseases or conditions associated
CC with inappropriate FGFR-5 expression and activity. For example, the
CC nucleic acids (and vectors containing them) and the FGFR-5 polypeptide
CC may be used to treat disorders associated with increased or decreased
CC cell survival (such as cancers (e.g. leukemia, colonic cancer,
CC testicular cancer and follicular lymphomas), autoimmune disorders (e.g.
CC multiple sclerosis and Crohn's disease) viral infections (e.g. herpes
CC viruses), inflammation, graft versus host disease, acute and chronic
CC graft rejection, ischemic injuries and atherosclerosis), activation,
CC secretion, migration, differentiation and proliferation, diseases
CC associated with defects in wound healing, mucosistosis, defects of
CC angiogenesis, immune dysfunction, endocrine dysfunction and insulin
CC secretion disorders. Anti-FGFR-5 antibodies may also be used as
CC diagnostic agents for detecting the presence of FGFR-5 polypeptides in
CC samples.

SO Sequence 46 BP; 11 A; 11 C; 19 G; 5 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 46;
Best Local Similarity 95.0%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtggccgagcagcagcagc 20
|||||
Db 17 gtggccgagcagcagcagc 36

RESULT 14

AAA28847
ID AAA28847 standard; DNA; 46 BP.

AC AAA28847;

DT 29-AUG-2000 (first entry)

DE 3' primer for full length extracellular domain FGFR5 DNA amplification.

KW FGFR-5: fibroblast growth factor receptor 5; cytosolic; anti-sclerotic;

KW immunomodulatory; gastrointestinal; virucide; anti-inflammatory;

KW anti-ischemic; anti-atherosclerosis; angiogenic; endocrine;

KW anti-diabetic; gene therapy; primer; ss.

OS Homo sapiens.

PN WO200024756-A1.

PD 04-MAY-2000.

PF 17-JUN-1999; 99WO-US13620.

PR 23-OCT-1998; 98US-0105465.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Young PE;

DR WPI; 2000-387035/33.

PT Nucleic acids encoding fibroblast growth factor-5 useful for the

PT prevention, diagnosis and treatment of conditions associated with

PT tissue repair and aberrant cell functions, e.g. cell survival and

PT proliferation

PS Example 7; Page 106; 182pp; English.

CC associated with inappropriate FGFR-5 expression and activity. For
CC example, the nucleic acids (and vectors containing them) and the FGFR-5
CC polypeptide may be used to treat disorders associated with increased or
CC decreased cell survival (such as cancers (e.g. leukemia, colonic cancer,
CC testicular cancer and follicular lymphomas), autoimmune disorders (e.g.
CC multiple sclerosis and Crohn's disease) viral infections (e.g. herpes
CC viruses), inflammation, graft versus host disease, acute and chronic
CC graft rejection, ischemic injuries and atherosclerosis), activation,
CC secretion, migration, differentiation and proliferation, diseases
CC associated with defects in wound healing, mucosistosis, defects of
CC angiogenesis, immune dysfunction, endocrine dysfunction and insulin
CC secretion disorders. Anti-FGFR-5 antibodies may also be used as
CC diagnostic agents for detecting the presence of FGFR-5 polypeptides in
CC samples.

SO Sequence 46 BP; 11 A; 11 C; 19 G; 5 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 46;
Best Local Similarity 95.0%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtggccgagcagcagcagc 20
|||||
Db 17 gtggccgagcagcagcagc 36

RESULT 15

AAH05404/C
ID AAH05404 standard; cDNA; 854 BP.

AC AAH05404;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:2239.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELT-) HELIX RES INSTR.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 1; SEQ ID 2239; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 854 BP; 194 A; 303 C; 233 G; 123 T; 1 other;

Query Match 92.0%; Score 18.4; DB 22; Length 854;
Best Local Similarity 95.0%; Pred. No. 85;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gtgcccagagcagagac 20
|||||
DB 52 GTGCCGACGACGACGACGC 33

Search completed: August 14, 2002, 22:06:20
Job time: 11675 sec

PT effects, for treating cancer -

XX

XX Disclosure; Page 99; 145pp; English.

XX

CC The present invention relates to a method of reducing the toxicity of

CC flavopiridol by administration in combination with a second agent that

CC increases conjugative enzyme activity or glucuronosyltransferase

CC activity. This second agent should be capable of inhibiting biliary

CC transport and may be a uridine 5'-diphosphate glucuronosyltransferase such

CC as UGT1A9. The method can be used in the treatment of cancer,

CC gastrointestinal diseases and parasitic diseases. The present sequence is

CC a PCR primer used to amplify SNPs in the UGT1A9 coding sequence.

XX

XX

SO Sequence 24 BP; 3 A; 7 C; 6 G; 8 T; 0 other;

Query Match 100.0%; Score 24; DB 24; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.036; Mismatches 0; Indels 0; Gaps 0;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24

Db 1 ctccgcctcagactgtttgtag 24

RESULT 2

AAV17224

ID AAV17224 standard; DNA; 355 BP.

XX

AC AAV17224;

XX

XX 29-JUN-1998 (first entry)

XX

XX SCA2 gene fragment.

XX

XX SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

XX

XX Synthetic.

XX

XX Key Location/Qualifiers

FT CDS 341..355

FT /*tag= a

FT /note= "SCA2 protein fragment"

XX

XX

XX MO9803679-A1.

XX

XX 29-JAN-1998.

XX

XX

XX 18-JUL-1996; 96MO-JP01999.

XX

XX 18-JUL-1996; 96MO-JP01999.

XX

XX (SRLS-) SRL INC.

XX

XX Sanpei K, Tsuji S;

XX

XX WPI; 1998-120796/11.

DR P-PSDB; AAW41370.

XX

XX

XX Diagnosing spinocerebellar ataxis type II - by PCR and determining

PT number of CAG repeat units

XX

XX

PS Claim 1; Page 10; 23pp; Japanese.

XX

XX This sequence represents a fragment of the SCA2 gene. It can be used in

CC the method of the invention for diagnosing spinocerebellar ataxis type

CC II, by performing PCR on the test DNA using two primers hybridizing to

CC parts of the SCA2 gene sequence, and determining the number of CAG

CC repeats in the amplified products. The method provides an easy means for

CC the diagnosis of spinocerebellar ataxis type II.

XX

XX

XX Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

Query Match 100.0%; Score 24; DB 19; Length 355;

Best Local Similarity 100.0%; Pred. No. 0.053;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24

Db 73 ctccgcctcagactgtttgtag 96

RESULT 3

AAV17229

ID AAV17229 standard; DNA; 623 BP.

XX

AC AAV17229;

XX

XX 29-JUN-1998 (first entry)

XX

XX SCA2 gene fragment.

XX

XX SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

XX

XX Synthetic.

XX

XX Key Location/Qualifiers

FT CDS 341..583

FT /*tag= a

FT /note= "SCA2 protein fragment, no stop codon given"

XX

XX

XX MO9803679-A1.

XX

XX 29-JAN-1998.

XX

XX

XX 18-JUL-1996; 96MO-JP01999.

XX

XX 18-JUL-1996; 96MO-JP01999.

XX

XX (SRLS-) SRL INC.

XX

XX Sanpei K, Tsuji S;

XX

XX WPI; 1998-120796/11.

DR P-PSDB; AAW41372.

XX

XX

XX Diagnosing spinocerebellar ataxis type II - by PCR and determining

PT number of CAG repeat units

XX

XX

PS Example 1; Page 11-12; 23pp; Japanese.

XX

XX This sequence represents a fragment of the SCA2 gene. It can be used in

CC the method of the invention for diagnosing spinocerebellar ataxis type

CC II, by performing PCR on the test DNA using two primers hybridizing to

CC parts of the SCA2 gene sequence, and determining the number of CAG

CC repeats in the amplified products. The method provides an easy means for

CC the diagnosis of spinocerebellar ataxis type II.

XX

XX

XX Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 100.0%; Score 24; DB 19; Length 623;

Best Local Similarity 100.0%; Pred. No. 0.058;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24

Db 73 ctccgcctcagactgtttgtag 96

RESULT 4

AAV30270

ID AAV30270 standard; DNA; 4367 BP.

XX

AC AAV30270;

```

XX 02-OCT-1998 (first entry)
DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
XX
XX Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KM CAG repeat; neurodegenerative disease; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 49..3990
FT /*tag= a
FT /product= "Spinocerebellar ataxia type 2 associated
FT repeat_region protein"
FT 544..612
FT /*tag= b
FT /note= "normal CAG repeat region; this is increased in
FT patients with SCA2"
FT repeat_unit 544..546
FT /*tag= c
XX
XX W09818920-A1.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-JP03946.
XX
XX 30-OCT-1996; 96JP-0304059.
XX
XX (SRLS-) SRL INC.
XX
XX Sanpei K, Tsuji S;
XX
XX WPI: 1998-272215/24.
XX
XX P-PSDB; AAW60213.
XX
XX Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
XX
XX Claim 1; Pages 13-22; 38pp; Japanese.
XX
XX This represents the sequence of a gene causative of spinocerebellar
XX ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
XX with SCA2, has a tri-nucleotide (CAG) repeat region which in the
XX expression product produces a polyglutamine sequence from Gln-166 to
XX Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
XX patients this number is increased to 35-100. Peptides encoded by nucleic
XX acid fragments (DNA or RNA) containing sequences from the SCA2 associated
XX gene, antibodies recognising the peptides and antisense nucleic acids
XX hybridising with the nucleic acid fragments can be used for the
XX investigation and diagnosis of SCA2. They can also be used for the
XX treatment of SCA2 by antisense therapy or gene therapy.
XX
XX Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;
SQ

```

```

Query Match 100.0%; Score 24; DB 19; Length 4367;
Best Local Similarity 100.0%; Pred. NO. 0.077;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctcgcctcagactgtttgtag 24
DB 261 ctcgcctcagactgtttgtag 284

```

```

RESULT 5
AAV06552
ID AAV06552 standard: cDNA; 4481 BP.
XX
AC AAV06552;
XX

```

```

DT 06-JUL-1998 (first entry)
DE Human SCA2 cDNA including CAG repeat region.
XX
XX SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
KM diagnosis; olivoponto-cerebellar atrophy; ss; ds.
XX
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 164..4101
FT /*tag= a
FT complement (631..648)
FT /*tag= b
FT /note= "primer SCA2-A binding site"
FT 740..757
FT /*tag= c
FT /note= "primer SCA2-B binding site"
FT 1070..1091
FT /*tag= d
FT /note= "primer SCA2-14B binding site"
FT 899..900
FT /*tag= e
FT /note= "predicted splice site"
FT 658..723
FT /*tag= f
FT /note= "CAG repeat region"
FT 658..660
FT /*tag= g
FT /note= "CAG repeat"
FT 661..663
FT /*tag= h
FT /note= "CAG repeat"
FT 664..666
FT /*tag= i
FT /note= "CAG repeat"
FT 667..669
FT /*tag= j
FT /note= "CAG repeat"
FT 670..672
FT /*tag= k
FT /note= "CAG repeat"
FT 673..675
FT /*tag= l
FT /note= "CAG repeat"
FT 676..678
FT /*tag= m
FT /note= "CAG repeat"
FT 679..681
FT /*tag= n
FT /note= "CAG repeat"
FT 685..687
FT /*tag= o
FT /note= "CAG repeat"
FT 688..690
FT /*tag= p
FT /note= "CAG repeat"
FT 691..693
FT /*tag= q
FT /note= "CAG repeat"
FT 694..696
FT /*tag= r
FT /note= "CAG repeat"
FT 700..702
FT /*tag= s
FT /note= "CAG repeat"
FT 703..705
FT /*tag= t
FT /note= "CAG repeat"
FT 706..708
FT /*tag= u
FT /note= "CAG repeat"
FT 709..711

```

FT		/tag=	V
FT		/note=	"CAG repeat"
FT	repeat_unit	712..714	
FT		/tag=	w
FT		/note=	"CAG repeat"
FT	repeat_unit	715..717	
FT		/tag=	x
FT		/note=	"CAG repeat"
FT	repeat_unit	718..720	
FT		/tag=	y
FT		/note=	"CAG repeat"
FT	repeat_unit	721..723	
FT		/tag=	z
FT		/note=	"CAG repeat"
XX			
PN	W09742314-A1.		
XX			
PD	13-NOV-1997.		
XX			
PE	08-MAY-1997;	97WO-US07725.	
XX			
PR	08-OCT-1996;	96US-0727084.	
PR	08-MAY-1996;	96US-0017388.	
PR	19-JUL-1996;	96US-0022207.	
XX			
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.		
XX			
PI	Pulst S;		
XX			
DR	WPI; 1998-086523/08.		
DR	P-PSDB; AAM33807.		
XX			
PT	Nucleic acids encoding human and mouse ataxin 2 - a product of the		
PT	sphincerebellar ataxia 2 gene, SCA2; useful in the diagnosis of		
PT	ataxia type 2		
XX			
XX	Claim 6; Page 52-58; 98pp; English.		
XX			
CC	This cDNA sequence corresponds to a novel SCA2 gene encoding a human		
CC	sphincerebellar ataxis-2 (SCA2) polypeptide, designated ataxin-2		
CC	(see AAM33807). A trisomy 21 foetal brain cDNA library and an adult		
CC	human frontal cortex cDNA library in lambda zapII were screened		
CC	with probes obtained by PCR amplification of plasmid AAPE5122B (see		
CC	AAV06351). PCR products were used to screen the human adult frontal		
CC	cortex library, and 5' clones were obtained by RT-PCR of placental		
CC	mRNAs. Overlapping clones was used to generate the composite 4481		
CC	bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic		
CC	or transcribed mRNA sequence in an individual having an expanded		
CC	CAG repeat at a location corresponding to the CAG repeat region of		
CC	the SCA2 gene. The presence of at least 13 CAG repeats above the		
CC	normal level (22, occasionally 23, repeats) is indicative of SCA2.		
CC	Primers (see AAV99640-41) amplifying at least this region are used		
CC	for diagnosis. Also claimed are kits for detecting mutations at		
CC	the SCA2 locus, antisense oligonucleotides, and transgenic animals		
CC	useful for studying the physiological roles of ataxin-2 and its		
CC	effect upon behaviour.		
XX			
SO	Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;		
Query Match	100.0%; Score 24; DB 19; Length 4481;		
Best Local Similarity	100.0%; Pred. No. 0.078;		
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1 ctccgcctcagactgtttgtag 24		
Db	375 ctccgcctcagactgtttgtag 398		
RESULT	6		
AAZ23428			
ID	AAZ23428 standard; DNA; 4481 BP.		
XX			

AC	AA223428;
XX	
DT	19-JAN-2000 (first entry)
XX	
DE	Human SCA2 DNA.
XX	
KW	Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KM	huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA6;
KW	atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KM	Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW	dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KX	neoplastic; malignant; autoimmune; fibrotic; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
CDS	163..4101
FT	/tag= a
FT	/product= "SCA2"
PX	
PN	MO9945944-AI.
PD	
PD	16-SEP-1999.
PX	
PF	11-MAR-1999; 99WO-USO5250.
PX	
PR	12-MAR-1998; 98US-.0041886.
PX	
PA	(BURN-) BURNHAM INST.
PX	
P1	Bredesen DE, Rabizadeh S;
XX	
DR	WPI; 1999-561617/47.
XX	
XX	P-PSDB; AAY33495.
PT	New proapoptotic dependence peptides, used to develop products for
PT	treating, e.g. Alzheimer's disease -
XX	
PS	Disclosure: Page 130-135; 1999P; English.
XX	
CC	This invention describes novel pure proapoptotic dependence peptides
CC	which comprise a sequence of an active dependence domain selected from
CC	dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC	huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC	SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC	of inducing cell death and can be used to develop products to mediate or
CC	inhibit apoptosis. The methods can be used for reducing the severity of
CC	a proapoptotic dependence domain mediated pathological conditions e.g.
CC	Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC	Spinocerebellar ataxias, dentatorubropallidoluysian atrophy,
CC	Machado-Joseph disease, stroke or head trauma. They can also be used for
CC	reducing the severity of a pathological condition mediated by upregulated
CC	cell proliferation or cell survival e.g. neoplastic, malignant,
CC	autoimmune or fibrotic conditions. This sequence encodes the human
CC	SCA2 polypeptide described in the method of the invention.
XX	
SQ	Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other:
Query Match	100.0%; Score 24; DB 20; Length 4481;
Best Local Similarity	100.0%; Pred. No. 0.078;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ctccgcctcagactgtttgtag 24
Db	375 ctccgcctcagactgtttgtag 398
RESULT 7	
ID	AAS59526
XX	AAS59526 standard; DNA; 20878 BP.
AC	AAS59526;

```

XX 13-FEB-2002 (first entry)
DT
XX
XX Propionibacterium acnes immunogenic protein encoding DNA #21.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant; ds.
XX
XX Propionibacterium acnes.
OS
XX
XX WO200181581-A2.
PN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR
XX 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX MPI; 2001-616774/71.
DR
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX
XX Claim 1: SEQ ID NO 21: 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and their
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
CC in infections of bone, joints and the central nervous system, however it
CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AA045516-AA045731, AA067497 and AA067498.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 20878 BP; 4447 A; 6432 C; 6132 G; 3865 T; 2 other:
SQ

```

```

Query Match          72.5%; Score 17.4; DB 23; Length 20878;
Best local Similarity 94.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 4 cgcctcagactgtttgt 22
    |||||
DB 5189 cgcctcagactgtttgt 5207

```

```

RESULT 8
ABL06830/c
ID ABL06830 standard; cDNA; 5742 BP.
XX

```

```

AC ABL06830;
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 14972.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
KM
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
PI P-PSDB; ABB62727.
XX
XX MPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
XX Claim 1: SEQ ID NO 14972: 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC sequences (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5742 BP; 1496 A; 1317 C; 1392 G; 1537 T; 0 other:
SQ

```

```

Query Match          71.7%; Score 17.2; DB 23; Length 5742;
Best local Similarity 86.4%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 ctcgcctcagactgtttgt 22
    |||||
DB 3001 CTCGCGCGCCCGACTGTTTGGT 2980

```

```

RESULT 9
ABAI7086/c
ID ABAI7086 standard; DNA; 30032 BP.
XX
XX ABAI7086;
AC
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Human nervous system related polynucleotide SEQ ID NO 9417.
DE
XX
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virocid;
KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KM antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KM antitumoric; hepatotropic; cerebroprotective; antiinflammatory;
KM antiallergic; antidiabetic; antidiarrhoeal; anticonvulsant; antifungal;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;

```

KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244221.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-541565/60.
 DR
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 PS
 XX Disclosure: SEQ ID NO 9417; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 30032 BP; 6224 A; 9088 C; 9280 G; 5440 T; 0 other;
 XX
 Query Match 71.7%; Score 17.2; DB 22; Length 30032;
 Best Local Similarity 86.4%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 2 ctcgcctcagactgtttgta 23
 ||||| ||||| ||||| ||||| |||||
 DB 24188 TCCGCTCAGACAGCTTTGTA 24167
 RESULT 10
 AAC11508
 ID AAC11508 standard; CDNA; 226 BP.
 XX
 AC AAC11508;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 15583.
 XX
 KW Human: 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dunas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 15583; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 XX Sequence 226 BP; 39 A; 72 C; 56 G; 57 T; 2 other;
 XX
 Query Match 69.2%; Score 16.6; DB 21; Length 226;
 Best Local Similarity 82.6%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 ctcgcctcagactgtttgta 23
 ||||| ||||| ||||| ||||| |||||
 DB 101 ctcgcctcagactgtttgta 123
 RESULT 11
 AA187048/C
 ID AA187048 standard; CDNA; 378 BP.
 XX
 AC AA187048;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 7108.
 XX
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 XX
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 XX
 DR P-PSDB; AAO07117.
 XX
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 PS Claim 1; SEQ ID NO 7108; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 378 BP; 91 A; 94 C; 106 G; 87 T; 0 other;

Query Match 69.2%; Score 16.6; DB 22; Length 378;

Best Local Similarity 82.6%; Pred. No. 1.9e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 tccgcctcagactgtttgtag 24

Db 115 TCCGCTGAGACTGTTTCTCG 93

RESULT 12

AAH13897/C

ID AAH13897 standard; cDNA; 1658 BP.

AAH13897;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:10908.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-018776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID 10908; 2537pp + CD ROW; English.

The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SO Sequence 1658 BP; 448 A; 369 C; 429 G; 412 T; 0 other;

Query Match 69.2%; Score 16.6; DB 22; Length 1658;

Best Local Similarity 82.6%; Pred. No. 2.4e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 tccgcctcagactgtttgtag 24

Db 1244 TCTGCTCAGACTGTTTGAAG 1222

RESULT 13

AAF22778

ID AAF22778 standard; cDNA; 715 BP.

AAF22778;

26-MAR-2001 (first entry)

Human prostate cancer associated antigen nucleotide sequence SEQ ID:357.

Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

cancer associated antigen; cytostatic; cancer vaccine; ss.

Homo sapiens.

WO200073801-A2.

07-DEC-2000.

26-MAY-2000; 2000MO-US14749.

28-MAY-1999; 99US-0136526.

10-SEP-1999; 99US-0153454.

(LUDW-) LUDWIG INST CANCER RES.

Obata Y;

WPI; 2001-025274/03.

Claim 50; Page 390-391; 799pp; English.

AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB33232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 715 BP; 204 A; 181 C; 159 G; 165 T; 6 other;

Query Match 68.3%; Score 16.4; DB 22; Length 715;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 7 ctcaacgtgttgtag 24
||| |
Db 257 ctcaacgtgttgtag 274

RESULT 14
AAK82275/C
ID AAK82275 standard; DNA; 2281 BP.
XX
AC AAK82275;
XX
DT 07-NOV-2001 (first entry)
XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37087.

KX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX

OS Homo sapiens.

PN WO200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0235327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure: SEQ ID NO 37086; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins, and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 2580 BP; 670 A; 579 C; 632 G; 699 T; 0 other;

Query Match 68.3%; Score 16.4; DB 22; Length 2580;
Best Local Similarity 94.4%; Pred. No. 3.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 7 ctacagctgtttgtag 24
|||||
Db 341 CTCAAACTGTTTGGTAG 324

Search completed: August 14, 2002, 22:06:17
Job time: 11672 sec

Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccgcctcagactgtttgttag 24
|||||
Db 73 CTCGCCCTCAGACTGTTTGTAG 96

RESULT 2
ARI53580 ARI59558 572 bp DNA linear PAT 17-OCT-2001
LOCUS
DEFINITION Sequence 18 from patent US 6251589.
ACCESSION ARI59558
VERSION ARI59558.1 GI:16222251
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 572)
Tsuij,S. and Saneel,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers
therefor
JOURNAL Patent: US 6251589-A 18-26-JUN-2001;
FEATURES Location/Qualifiers
source 1..572
/organism="unknown"

BASE COUNT 34 a 277 c 174 g 85 t 2 others
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 572;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccgcctcagactgtttgttag 24
|||||
Db 73 CTCGCCCTCAGACTGTTTGTAG 96

RESULT 3
ARI59546 ARI59546 623 bp DNA linear PAT 17-OCT-2001
LOCUS
DEFINITION Sequence 5 from patent US 6251589.
ACCESSION ARI59546
VERSION ARI59546.1 GI:16222229
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 623)
Tsuij,S. and Saneel,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers
therefor
JOURNAL Patent: US 6251589-A 5-26-JUN-2001;
FEATURES Location/Qualifiers
source 1..623
/organism="unknown"

BASE COUNT 55 a 292 c 189 g 85 t 2 others
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccgcctcagactgtttgttag 24
|||||
Db 73 CTCGCCCTCAGACTGTTTGTAG 96

RESULT 4
ARI53580 ARI53580 4481 bp DNA linear PAT 08-AUG-2001
LOCUS

DEFINITION Sequence 18 from patent US 6235872.
ACCESSION ARI53580
VERSION ARI53580.1 GI:15121112
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 4481)
Bredesen,D.E. and Rabitzadeh,S.
TITLE Prapoptotic peptides dependence polypeptides and methods of use
therefor
JOURNAL Patent: US 6235872-A 18-22-MAY-2001;
FEATURES Location/Qualifiers
source 1..4481
/organism="unknown"

BASE COUNT 1144 a 1380 c 1014 g 943 t
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccgcctcagactgtttgttag 24
|||||
Db 375 CTCGCCCTCAGACTGTTTGTAG 398

RESULT 5
HSU70323 HSD70323 4481 bp mRNA linear PRI 20-NOV-1996
LOCUS
DEFINITION Human ataxin-2 (SCA2) mRNA, complete cds.
U70323
VERSION U70323.1 GI:1679683
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4481)
Pulst,S.-M., Nechiporuk,A., Nechiporuk,T., Gispert,S., Chen,X.-N.,
Lunkes,A., DeJong,P., Rouleau,G.A., Auburger,G., Korenberg,J.R.,
Figueroa,C. and Sahba,S.
TITLE Moderate expansion of a normally biallelic trinucleotide repeat in
spinocerebellar ataxia type 2
JOURNAL Nature Genet. 14 (3), 269-276 (1996)
MEDLINE 97051920
AUTHORS 2 (bases 1 to 4481)
Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1996) Medicine, Cedars-Sinai, 8700 Beverly Blvd.,
Los Angeles, CA 90048, USA
FEATURES Location/Qualifiers
source 1..4481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24.1"
1..4481
/gene="SCA2"
163..4101
/gene="SCA2"
/standard_name="spinocerebellar ataxia type 2"
/codon_start=1
/product="ataxin-2"
/protein_id="AAB19200.1"
/db_xref="GI:1679684"
/translation="MRSAAAPRSPAVATSRRAARWPCWRSLSQPARSGRGCGG
AAPGYPSPAPPPPGPPSPSSDCFGSNGNGGAGAFRGSRRLLGLGPPR
PFVVVLLPLASGAPPAATTASPLGARSPRSGLARAPGCPRPACPPVYGPLT
MSLKP000000000000000000PPPAANYRKPGSGLLASPAAPSPSSSV
SSSAAPSPSSVAATSGGRRPLGGRSNGTLPOSTISPGIYANMVMVILITSVG
SKCEVQVANKGIVGIVFKYIPKCDLVDAHAHEKTESSSGPKKEIMESTLIFKCSDF

VVGFDMDSYAKRDAFTDSAISAKVNGEHEKEDLEPMAGELTANBELALENDVS
 NGWDMDFREYENGYVSTYDSSLSSTYVLEBDNSEEFKLRARNOQLAELESS
 AOYKARVALENDDBSEFEKYTAQVONSSRECHSINTRENYKIPPCONRREISMCSG
 RONSFRMGQSGSGSPRSSTSTHTDFNPNSSGDQRYVNGCVWPEPCSPSPRSRY
 OSGPNLPPRAATPPRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSR
 KAORHRHNRVSAAGSISISGLEFVSHNPSEAPVARSPPSGTWSVSGVPL
 SPKTRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSRPSR
 PSSEKDSRLQDORONSPAGNENIKPNPSTSPSKAEKNGISPVYSEKROIDLCK
 FKPNFRLOPSSSTPSEMDOLNKNRECEKSRDILKIPESAKDSFLENSMCTGSS
 KNPSPSISSTISLNTREHKGPEVTSOGVOTSSPACKOKEKDKEKDAEYKSTLN
 PNAKEFNPRSEFQPKPSTTPTSPRPOAOCSFPMVGHQOFTPTVTPCPCAPMMTPVP
 VSPGVOLPIPIPTMPVNAQTYRAVPMQRODHOHOSAMHPASAAPPATP
 PAYSTOYVSPQEPNPOLVOHVPYHOSQHPVSPVOCNARMAPPTHPAOLVS
 SSATQYGAHEOTHAAYACPKLYNKETSPSEFALSTGSLAOQVHPNATLPHPTHP
 OPSATPGOOSGOSGSHAPSPVPOHOOAOLHLASPOOASLYHAGLAPTPPSM
 TPASNTOSRONSFPAAOQTVTFIHSHPQATNPNNPHANVQAHVOSMVSHTPLAH
 APWMLTTPQPGPQALAQSLADPLPSTTAHPFMTHTPSQAHHQOOL*

BASE COUNT 1144 a 1380 c 1014 g 943 t
 ORIGIN

Query Match 100.0%; Score 24; DB 9; Length 4481;
 Best Local Similarity 100.0%; Pred. No. 0.066;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcgcctcagactgtttgtag 24
 Db 375 CTCGCCCTCAGACTGTTTGTAG 398

RESULT 6
 AC004085/231758 bp DNA linear HTG 06-NOV-2000
 LOCUS Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered
 DEFINITION
 AC004085
 AC004085.6 GI:11079383
 VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Plimates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 231758)
 Muzny,D.M., Adams,C., Adio,-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alstrooms,S.L., Amartunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burgett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Din,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gablis,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
 Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L.J., Kovav,J.,
 Kover,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
 Moran,M., Morris,S., Moser,M., Neal,D., Newtonson,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwko,S.,
 Ogub,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Kojubokan,I., Rolfe,M.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Rule,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tametisa,A., Tametisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Washington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,
 Morley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.
 and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 231758)
 Morley,K.C.
 Direct Submission
 Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 3, 2000 this sequence version replaced gi:9966929.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center Project name: UG
 Center clone name: RP11-42B1
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 224788 bases at least Q40
 Consensus quality: 229074 bases at least Q30
 Consensus quality: 230948 bases at least Q20
 Estimated insert size: 227237; sum-of-contigs estimation
 Estimated insert size: 317311; agarose-gel estimation
 Quality coverage: 6.3x in Q20 bases; agarose-gel estimation
 Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 33241: contig of 33241 bp in length
 * 33242 33241: gap of unknown length
 * 33342 56391: contig of 23050 bp in length
 * 56392 81333: gap of unknown length
 * 56492 81333: contig of 24832 bp in length
 * 81324 81423: gap of unknown length
 * 81424 102538: contig of 21155 bp in length
 * 102539 102638: gap of unknown length
 * 102639 119710: contig of 17072 bp in length
 * 119711 119810: gap of unknown length
 * 119811 136913: contig of 17103 bp in length
 * 136914 137014: gap of unknown length
 * 137014 153285: contig of 16272 bp in length
 * 153286 153385: gap of unknown length
 * 153386 167987: contig of 14602 bp in length
 * 167988 168087: gap of unknown length
 * 168088 178731: contig of 10644 bp in length
 * 178732 178831: gap of unknown length
 * 178832 186641: contig of 7810 bp in length
 * 186642 186741: gap of unknown length
 * 186742 193215: contig of 6474 bp in length
 * 193216 193315: gap of unknown length
 * 193316 201310: contig of 7995 bp in length
 * 201311 208647: contig of 7237 bp in length
 * 208648 208747: gap of unknown length
 * 208748 213802: contig of 5055 bp in length
 * 213802 213902: gap of unknown length

```

* 213903 218049: contig of 4147 bp in length
* 218050 218149: gap of unknown length
* 218150 223316: contig of 5167 bp in length
* 223317 223416: gap of unknown length
* 223417 227389: contig of 3973 bp in length
* 227390 227489: gap of unknown length
* 227490 229033: contig of 1543 bp in length
* 229033 230651: gap of unknown length
* 230652 230751: gap of unknown length
* 230752 231758: contig of 1007 bp in length.
Location/Qualifiers
1..231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"

```

```

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

```

```

Query Match 100.0%; Score 24; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ctccgcctcagactgtttgtag 24
|||||
Db 89411 CTCCGCTCAGACTGTGTGTAG 89388

```

```

RESULT 7
LOCUS AF041472 4225 bp mRNA linear ROD 28-NOV-2001
DEFINITION Mus musculus ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION AF041472
VERSION AF041472.1 GI:3005019
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4225)
Nechiporuk,T.T., Huynh,D.P., Figueroa,K., Sabha,S., Nechiporuk,A.V.
and Pulst,S.M.
The mouse SCA2 gene: cDNA sequence, alternative splicing and
protein expression.
JOURNAL Hum. Mol. Genet. 7 (8), 1301-1309 (1998)
MEDLINE 9668173
PUBMED 98334550

```

```

TITLE 2 (bases 1 to 4225)
Nechiporuk,T.T., Figueroa,K., Sabha,S., Nechiporuk,A.V. and
Pulst,S.M.
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Medicine/Neurology, Cedars-Sinai Medical
Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA

```

FEATURES

```

source
1..4225
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="12q23.1"
1..4225
/gene="SCA2"
27..3884
/gene="SCA2"
/codon_start=1
/product="ataxin-2"
/protein_id="AAC09275.1"
/db_xref="GI:3005020"
/translation="MRSSTAAVQRPADGPPRRPAGMAARSLPRTARAGRGGAVA
TPSAGPPRRGAPGRRSPGASDGFSGNGHGSRRSLRLGCGPPRFVYVL
ALAPRAATPARACPGVRASPPRGVSSASAPAPGCPACPPYGLMLSKPOOP
APATGRKPGGGLSSPGAAPASAAVTSVYPAAPAPYASSSAAAGGRPGICGRNS
SKGLPPTISFDGIYANVMWHLITSVGSKCEVQKNGIIEGVFKTYSKCDLVLD

```

```

AAHEKSTESSGPKREIMESVLFKCSDFVVOVKDTSYARARAFDTSALSAKNG
EHKEKLEPMDAGELTASBELLENDVSNQNDPFRVNEENVGYSTYDSLSYLT
VPLERDNSEEFLEKREARANOIAEIESAOYKARVALENDRESEKRYAVORNCSDR
EGHGPTNRDKYTPGQRNREVSVMGSGRSPRMOPGSGMPSRAASITSPFNNA
GSDQRYVNGVYPMPPSPCPSHSRPSYSGRPSLPRAATHRPSRPSRPSRPS
HPSAHSAPAVSTMPKMSSEGGPPRSPKORHPRNRHSAGCSMSGLERYSHNP
SEAAAPVARTSPAGCTSSVSVGNALTPSIEAKDSRLDQQRONSAGSKENYKASST
IIPAEAVSMVPAPASPPAPASGNALTPSIEAKDSRLDQQRONSAGSKENYKASST
SPFSKADNKGMSPVVSEHRKQIDDKFKFNDRLOPSSYSESDOLLSKNRGEKSR
DLIKDTEASAKDSFIDSSSSSCKTSGSKTNSPISPSMLNAEHRKGEVTSQGV
OTSSPACOKRDKDRBEKKDTTEQVRSKTLNPNKKEPNSFSQPKSTPTSPRPAO
PSPSMOCHOPAPVYTPQPCFARNMYPVPSGVOPLPIPMTPMVPMDAKTYRAGK
VPMPOORQDHOHSTMMHPASAGPPTVATTPRATISQYAIYSPOQPPNOLVQHPH
YOSQHPHVSPIVIOGNARMAPPAHAQGLVSSAOPCAHEQTHMATICPKLPTNKE
TSPSEFALISTGSLAQOYAHNMAALHPHTBPQSAITPTGQOOSQHGSGHPVQV
HOHOAOLALHMASPOOASAIYHAGLAFTPTSPMPSASTOSPOSFPAADQVFTIHS
HVOAPATTPPHMAHVPOAHVOSGVPSHPTAHAPMMLMTQTPGPKRAALAAQSALQPIP
VSTTAHPRVYTHPSVOAHHQOOL"

```

```

BASE COUNT 1007 a 1324 c 1042 g 851 t 1 others
ORIGIN

```

```

Query Match 91.7%; Score 22; DB 10; Length 4225;
Best Local Similarity 91.7%; Pred. No. 0.8;
Matches 22; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 ctccgcctcagactgtttgtag 24
|||||
Db 224 STGCCCTCAGACTGTGTGTAG 247

```

```

RESULT 8
LOCUS AC010803/c 146487 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone Rpl1-2024, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC010803
VERSION AC010803.3 GI:9120116
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 146487)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
JOURNAL 2 (bases 1 to 146487)

```

```

TITLE 2 (bases 1 to 146487)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
AUTHORS

```

```

TITLE Unpublished
JOURNAL 2 (bases 1 to 146487)

```

```

REFERENCE 1 (bases 1 to 146487)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
JOURNAL 2 (bases 1 to 146487)
MEDLINE 146487
PUBMED 146487

```

```

TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

COMMENT All repeats were identified using RepeatMasker.
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L2687
Center clone name: 2_O_24

* NOTE: This record contains 148 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 866: contig of 866 bp in length
867 966: gap of 100 bp
967 1831: contig of 865 bp in length
1832 1931: gap of 100 bp
1932 2814: contig of 883 bp in length
2815 2914: gap of 100 bp
2915 3793: contig of 879 bp in length
3794 3893: gap of 100 bp
3894 4777: contig of 884 bp in length
4778 4877: gap of 100 bp
4878 5733: contig of 856 bp in length
5734 5833: gap of 100 bp
5834 6709: contig of 876 bp in length
6710 6809: gap of 100 bp
6810 7681: contig of 872 bp in length
7682 7781: gap of 100 bp
7782 8662: contig of 881 bp in length
8663 8762: gap of 100 bp
8763 9643: contig of 881 bp in length
9644 9743: gap of 100 bp
9744 10613: contig of 870 bp in length
10614 10713: gap of 100 bp
10714 11570: contig of 857 bp in length
11571 11670: gap of 100 bp
11671 12555: contig of 885 bp in length
12556 12655: gap of 100 bp
12656 13508: contig of 853 bp in length
13509 13608: gap of 100 bp
13609 14490: contig of 882 bp in length
14491 14590: gap of 100 bp
14591 15479: contig of 889 bp in length
15480 15579: gap of 100 bp
15580 16444: contig of 865 bp in length
16445 16544: gap of 100 bp
16545 17426: contig of 882 bp in length
17427 17526: gap of 100 bp
17527 18399: contig of 873 bp in length
18400 18499: gap of 100 bp
18500 19392: contig of 893 bp in length
19393 19492: gap of 100 bp
19493 20385: contig of 893 bp in length
20386 20485: gap of 100 bp
20486 21361: contig of 876 bp in length
21362 21461: gap of 100 bp
21462 22337: contig of 876 bp in length
22338 22437: gap of 100 bp
22438 23319: contig of 882 bp in length
23320 23419: gap of 100 bp
23420 24296: contig of 877 bp in length
24297 24396: gap of 100 bp
24397 25255: contig of 859 bp in length
25256 25355: gap of 100 bp
25356 26264: contig of 909 bp in length
26265 26364: gap of 100 bp
26365 27236: contig of 872 bp in length
27237 27336: gap of 100 bp
27337 28201: contig of 865 bp in length
28202 28301: gap of 100 bp

28302 29192: contig of 891 bp in length
29193 29292: gap of 100 bp
29293 30161: contig of 869 bp in length
30162 30261: gap of 100 bp
30262 31139: contig of 878 bp in length
31140 31239: gap of 100 bp
31240 31240: contig of 874 bp in length
31241 32213: gap of 100 bp
32214 33083: contig of 870 bp in length
33084 33183: gap of 100 bp
33184 34060: contig of 877 bp in length
34061 34160: gap of 100 bp
34161 35048: contig of 888 bp in length
35049 35148: gap of 100 bp
35149 36024: contig of 876 bp in length
36025 36124: gap of 100 bp
36125 36979: contig of 855 bp in length
36980 37079: gap of 100 bp
37080 37941: contig of 862 bp in length
37942 38041: gap of 100 bp
38042 38937: contig of 886 bp in length
38938 39037: gap of 100 bp
39038 39916: contig of 879 bp in length
39917 40016: gap of 100 bp
40017 40899: contig of 883 bp in length
40900 40999: gap of 100 bp
41000 41913: contig of 914 bp in length
41914 42013: gap of 100 bp
42014 42901: contig of 888 bp in length
42902 43001: gap of 100 bp
43002 43868: contig of 867 bp in length
43869 43968: gap of 100 bp
43969 44824: contig of 856 bp in length
44825 44924: gap of 100 bp
44925 45793: contig of 869 bp in length
45794 45893: gap of 100 bp
45894 46741: contig of 848 bp in length
46742 46841: gap of 100 bp
46842 47736: contig of 895 bp in length
47737 47836: gap of 100 bp
47837 48716: contig of 880 bp in length
48717 48816: gap of 100 bp
48817 49686: contig of 870 bp in length
49687 49786: gap of 100 bp
49787 50658: contig of 872 bp in length
50659 50758: gap of 100 bp
50759 51634: contig of 876 bp in length
51635 51734: gap of 100 bp
51735 52630: contig of 896 bp in length
52631 52730: gap of 100 bp
52731 53619: contig of 889 bp in length
53620 53719: gap of 100 bp
53720 54578: contig of 859 bp in length
54579 54678: gap of 100 bp
54679 55554: contig of 876 bp in length
55555 55654: gap of 100 bp
55655 56547: contig of 893 bp in length
56548 56647: gap of 100 bp
56648 57532: contig of 885 bp in length
57533 57632: gap of 100 bp
57633 58506: contig of 874 bp in length
58507 58606: gap of 100 bp
58607 59505: contig of 899 bp in length
59506 59605: gap of 100 bp
59606 60506: contig of 901 bp in length
60507 60606: gap of 100 bp
60607 61492: contig of 886 bp in length
61493 61592: gap of 100 bp
61593 62514: contig of 922 bp in length
62515 62614: gap of 100 bp
62615 63498: contig of 884 bp in length
63499 63598: gap of 100 bp
63599 64483: contig of 885 bp in length

* 64484 64583: gap of 100 bp
* 64584 65452: contig of 869 bp in length
* 65453 65552: gap of 100 bp
* 65553 66437: contig of 885 bp in length
* 66438 66537: gap of 100 bp
* 66538 67414: contig of 877 bp in length
* 67415 67514: gap of 100 bp
* 67515 68393: contig of 879 bp in length
* 68394 68493: gap of 100 bp
* 68494 69344: contig of 851 bp in length
* 69345 69444: gap of 100 bp
* 69445 70299: contig of 855 bp in length
* 70300 70399: gap of 100 bp
* 70400 71285: contig of 886 bp in length

Query Match 75.8%; Score 18.2; DB 2; Length 146487;
Best Local Similarity 87.0%; Pred. No. 80;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cccgcctcagactgtttgtag 23
Db 146058 CTCCTCCACACGTTTGTGTA 146036
|||||

RESULT 9
LOCUS AC068040 182035 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-643022, WORKING DRAFT
ACCESSION AC068040
VERSION AC068040.7 GI:14626405
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 182035)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 182035)
Waterston, R.H.
Direct Submission
Submitted (27-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2001 this sequence version replaced gi:11545982.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0643022

----- Summary Statistics -----
Sequencing vector: M13: 568
Sequencing vector: plasmid: 438
Chemistry: Dye-terminator ET: 568 of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180296 bases at least Q40
Consensus quality: 180748 bases at least Q30
Insert size: 187000; agarose-fp
Quality coverage: 10.30 in Q20 bases; agarose-fp
Quality coverage: 10.12 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* Runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 13717: contig of 13717 bp in length
* 13718 13817: gap of unknown length
* 13818 15832: contig of 2015 bp in length
* 15833 15932: gap of unknown length
* 15933 73978: contig of 58046 bp in length
* 73979 74078: gap of unknown length
* 74079 182035: contig of 107957 bp in length.
Location/Qualifiers
1. 182035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-643022"
1. 13717
/note="assembly-name:Contig4
clone-end:77
vector_side:left"
13818. 15832
/note="assembly-name:Contig3"
15933. 73978
/note="assembly-name:Contig5"
74079. 182035
/note="assembly-name:Contig6"
BASE COUNT 51741 a 41726 c 41192 g 47076 t 300 others
ORIGIN

Query Match 75.8%; Score 18.2; DB 2; Length 182035;
Best Local Similarity 87.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 tccgcctcagactgtttgtag 24
Db 70085 TCCTCCACACGTTTGTGTA 70063
|||||

RESULT 10
LOCUS AC016906 183045 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-436E24 from 2, complete sequence.
ACCESSION AC016906
VERSION AC016906.7 GI:14589732
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183045)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
2 (bases 1 to 183045)
Shah, N. and Meyer, R.
The sequence of Homo sapiens BAC clone RP11-436E24
Unpublished (2001)
3 (bases 1 to 183045)
Waterston, R.H.
Direct Submission
Submitted (08-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 183045)
Waterston, R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 183045)


```

repeat_region 21920..22286 /rpt_family="MaLR"
misc_feature 22096..22607 /note="similar to EST AI884694 (NTD:95589858) w183e04.x1"
repeat_region 22462..22926 /rpt_family="MERL_type"
repeat_region 22962..23117 /rpt_family="MERL_type"
repeat_region 23366..23393 /rpt_family="MERL_type"
repeat_region 23394..23577 /rpt_family="Achoho"
repeat_region 23452..23475 /rpt_family="AT-rich"
repeat_region 23596..23713 /rpt_family="MIR"
repeat_region 24516..25608 /rpt_family="L2"
repeat_region 26170..26271 /rpt_family="MaLR"
repeat_region 26272..26751 /rpt_family="MaLR"
repeat_region 26752..27106 /rpt_family="MaLR"

Query Match 75.8%; Score 18.2; DB 9; Length 183045;
Best Local Similarity 87.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tccgcctcagactgtttgtag 24
Db 87089 TGCTCCTCAGACTGTCTGTGTAG 87111

RESULT 11
AC009285 183155 bp DNA linear HTG 22-DEC-2001
LOCUS Homo sapiens chromosome 18 clone RP11-186J15 map 18, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC009285 GI:17977456
VERSION AC009285.6
KEYWORDS HTG; HTGS-PHASE2; HTGS-FULLTOP; HTGS-ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183155)
Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
Baker J., Baldwin J., Barna N., Beckerly R., Benn J., Brown A.,
Castle J., Cerny J., Colangelo M., Collins S., Collymore A.,
Cooke P., Dearellano K., Depayre E., Devon K., Dewar K.,
Donegan L., Doyle M., Ferreira P., FitzHugh W., Forrest C.,
Funke R., Gage D., Galagan J., Gardyna S., Gilbert D., Grant G.,
Hagos B., Heaford A., Horton L., Howland J.C., Jones C., Kann L.,
Karatas A., Lehoczy J., Lien C., Locke K., Macdonald P.,
Marquis N., McEwan P., McGurk A., McKernan K., McLaughlin J.,
Meldrum J., Molla M., Morris W., Morrice J., Mychaleckyj J.,
Naylor J., Niloff M., O'Connor F., O'Donnell P., Pavlin B.,
Peterson K., Pollara V., Riley R., Roberts D., Roy A., Severy P.,
Stange-Thomann N., Stojanovic N., Stone C., Subramanian A.,
Tesfaye S., Torruella-Miller I., Vassiliev H., Vo A., Wagner A.,
Wheeler J., Wu X., Wyman D., Ye W.J. and Zody M.
Direct Submission
Submitted (12-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 22, 2001 this sequence version replaced g1:15487431.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1982
Center clone name: 186-J_15
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 156822: contig of 156822 bp in length
* 156823 156922: gap of 100 bp
* 156923 183155: contig of 26233 bp in length.
* Location/Qualifiers
  1..183155
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="18"
    /map="18"
    /clone="RP11-186J15"
    /clone_11b="RP11-186J15" Human Male BAC"
BASE COUNT 58917 a 35380 c 33020 g 55700 t 138 others
ORIGIN

Query Match 75.8%; Score 18.2; DB 2; Length 183155;
Best Local Similarity 87.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tccgcctcagactgtttgtag 24
Db 106403 TCCGCCTCAGACTGTCTGTGTAG 106381

RESULT 12
AC061962
ID AC061962 standard; DNA; HTG; 197626 BP.
XX AC061962;
AC AC061962.3
XX
XX 26-APR-2000 (Rel. 63, Created)
DT 01-JUL-2000 (Rel. 64, last updated, Version 3)
XX
XX Homo sapiens chromosome 2 clone RP11-764D5 map 2, WORKING DRAFT SEQUENCE,
DE 23 unordered pieces.
XX
XX HTG; HTGS-DRAFT; HTGS-PHASE1.
XX
XX Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
XX
XX [1]
XX 1-197626
XX Birren B., Linton L., Nusbaum C., Lander E.,
XX 1-197626
XX Birren B., Linton L., Nusbaum C., Lander E.,
XX "Homo sapiens chromosome 2, clone RP11-764D5";
XX Unpublished.
XX
XX [2]
XX 1-197626
XX Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
XX 1-197626
XX Birren B., Baldwin J., Barna N., Bastien V., Beda F., Boguslavsky L.,
XX Boukhalter B., Brown A., Burkett G., Campolano A., Castle A., Choedel Y.,
XX Colangelo M., Collins S., Collymore A., Cooke P., Dearellano K., Dewar K.,

```

RA Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., Fitzhugh W.,
RA Gage D., Galagan J., Gardyna S., Glinde S., Goyette M., Graham L.,
RA Grand-Pierre N., Grant G., Hagos B., Hearford A., Horton L., Howland J.C.,
RA Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., Lacombe K.,
RA Lamas R., Landers T., Lenczky J., Levine R., Lieu C., Liu G.,
RA Locke K., Macdonald P., Marquis N., McCarthy M., McEwan P., McGuirk A.,
RA McEwan K., McPeckers R., Meldrum J., Menus L., Mihova T., Miranda C.,
RA Mianga V., Morrow J., Murphy T., Naylor J., Norman C.H., O'Connor T.,
RA O'Donnell P., O'Neil D., Olivier T.M., Oliver J., Peterson K., Pierre N.,
RA Pisanil C., Pollata V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,
RA Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,
RA Stojanovic N., Subramanian A., Talamas J., Testaye S., Theodore J.,
RA Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A.,
RA Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A.,
Zody M.;
; Submitted (21-APR-2000) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
CC On Jun 21, 2000 this sequence version replaced gi:8135816.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: MIBR
CC Web site: <http://www-seq.wi.mit.edu>
CC Contact: sequence_submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L10013
CC Center clone name: 764_D_5
CC ----- Summary Statistics
CC Sequencing vector: M13; M77815; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 184877 bases at least Q40
CC Consensus quality: 191407 bases at least Q30
CC Consensus quality: 194008 bases at least Q20
CC Insert size: 212000; agarose-fp
CC Insert size: 195426; sum-of-ctrls
CC Quality coverage: 4.6 in Q20 bases; agarose-fp
CC Quality coverage: 5.0 in Q20 bases; sum-of-ctrls
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 23 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC *
CC 1 1987 2086: gap of 1986 bp in length
CC *
CC 2087 3545: contig of 1459 bp in length
CC *
CC 3546 3645: gap of 100 bp
CC *
CC 3646 6028: contig of 2383 bp in length
CC *
CC 6029 6128: gap of 100 bp
CC *
CC 6129 10083: contig of 3955 bp in length
CC *
CC 10084 10183: gap of 100 bp
CC *
CC 10184 13113: contig of 2930 bp in length
CC *
CC 13114 13213: gap of 100 bp
CC *
CC 13214 16896: contig of 3683 bp in length
CC *
CC 16897 16996: gap of 100 bp
CC *
CC 16997 19524: contig of 2528 bp in length
CC *
CC 19525 19624: gap of 100 bp
CC *
CC 19625 22626: contig of 3002 bp in length
CC *
CC 22627 22726: gap of 100 bp
CC *
CC 22727 28466: contig of 5740 bp in length
CC *
CC 28467 28566: gap of 100 bp
CC *
CC 28567 32304: contig of 3738 bp in length
CC *
CC 32305 32404: gap of 100 bp
CC *
CC 32405 37212: contig of 4808 bp in length

CC * 37213 37312: gap of 100 bp
CC * 37313 43344: contig of 6032 bp in length
CC * 43345 43444: gap of 100 bp
CC * 43445 47313: contig of 3869 bp in length
CC * 47314 47413: gap of 100 bp
CC * 47414 54652: contig of 7239 bp in length
CC * 54653 54752: gap of 100 bp
CC * 54753 68388: contig of 13636 bp in length
CC * 68389 68488: gap of 100 bp
CC * 68489 83530: contig of 15042 bp in length
CC * 83531 83630: gap of 100 bp
CC * 83631 93424: contig of 9794 bp in length
CC * 93425 93524: gap of 100 bp
CC * 93425 108652: contig of 15128 bp in length
CC * 108653 108752: gap of 100 bp
CC * 108753 121355: contig of 12603 bp in length
CC * 121356 121455: gap of 100 bp
CC * 121456 137482: contig of 16027 bp in length
CC * 137483 137582: gap of 100 bp
CC * 137583 152217: contig of 14635 bp in length
CC * 152218 152317: gap of 100 bp
CC * 152318 170994: contig of 18677 bp in length
CC * 170995 171094: gap of 100 bp
CC * 171095 197626: contig of 26532 bp in length.
XX
FH Key Location/Qualifiers
FH
FH source
FT 1. 197626
FT /chromosome="2"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /map="2"
FT /clone="RP11-764D5"
FT /clone_lib="RC1-11 Human Male BAC"
FT 1. 1986
FT /note="assembly-fragment"
FT 2087. 3545
FT /note="assembly-fragment clone_end:T7 vector_side:left"
FT 3546. 6028
FT /note="assembly-fragment"
FT 6129. 10083
FT /note="assembly-fragment"
FT 10184. 13113
FT /note="assembly-fragment"
FT 13214. 16896
FT /note="assembly-fragment"
FT 16997. 19524
FT /note="assembly-fragment"
FT 19625. 22626
FT /note="assembly-fragment"
FT 22727. 28466
FT /note="assembly-fragment"
FT 28567. 32304
FT /note="assembly-fragment"
FT 32405. 37212
FT /note="assembly-fragment"
FT 37313. 43344
FT /note="assembly-fragment"
FT 43445. 47313
FT /note="assembly-fragment clone_end:SP6 vector_side:left"
FT 47414. 54652
FT /note="assembly-fragment"
FT 54753. 68388
FT /note="assembly-fragment"
FT 68489. 83530
FT /note="assembly-fragment"
FT 83631. 93424
FT /note="assembly-fragment"
FT 93525. 108652
FT /note="assembly-fragment"
FT 108753. 121355
FT /note="assembly-fragment"
FT 121456. 137482
FT misc_feature

* This record will be updated with the finished sequence

4 (bases 1 to 221961)

TITLE Direct Submission
JOURNAL Submitted (29-JAN-2002) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Dec 24, 2001 this sequence version replaced gi:16756253.
----- Genome Center
Center: Department of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR

FEATURES
source Location/Qualifiers
1. 221961
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="rp23-383h20"
/clone_1lb="RP23"
BASE COUNT 66521 a 49019 c 46653 g 59767 t 1 others
ORIGIN
Query Match 75.8%; Score 18.2; DB 10; Length 221961;
Best Local Similarity 87.0%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 tccgcctcacagctgtttgtag 24
||| ||||| ||||| ||||| ||
Db 194891 TCCTCCTCAACGTCTTTGGGAG 194913

RESULT 15
AC094002/c 158023 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-156H20, *** SEQUENCING IN PROGRESS
DEFINITION *** 55 unordered pieces.
ACCESSION AC094002
VERSION AC094002.4 GI:17969845
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 158023)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alshrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Blinze,K., Blankenburg,K., Bonito,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burich,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Drepper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudon,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwoko,S.,
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158023)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Dec 20, 2001 this sequence version replaced gi:17062361.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCID
Center clone name: CH230-156H20
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findhaplplast
Consensus quality: 136844 bases at least Q40
Consensus quality: 142765 bases at least Q30
Consensus quality: 148321 bases at least Q20
Estimated insert size: 140986; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 8978: contig of 8978 bp in length
* 8979
* 9078: gap of unknown length
* 9079
* 15063: contig of 5985 bp in length
* 15064
* 15163: gap of unknown length
* 15164
* 23474: contig of 8311 bp in length
* 23475
* 23574: gap of unknown length
* 23575
* 30155: contig of 6581 bp in length
* 30156
* 30255: gap of unknown length
* 30256
* 36758: gap of 6503 bp in length
* 36759
* 36858: gap of unknown length
* 36859
* 41368: contig of 4510 bp in length
* 41369
* 41468: gap of unknown length
* 41469
* 45053: contig of 3585 bp in length
* 45054
* 45153: gap of unknown length
* 45154
* 50419: contig of 5266 bp in length
* 50420
* 50519: gap of unknown length
* 50520
* 52829: contig of 2310 bp in length
* 52830
* 52929: gap of unknown length
* 52930
* 56841: contig of 3912 bp in length
* 56842
* 56941: gap of unknown length
* 56942
* 60054: contig of 3113 bp in length
* 60055
* 60154: gap of unknown length
* 60155
* 65098: contig of 4944 bp in length
* 65198
* 65199: gap of unknown length
* 65199
* 68438: contig of 3240 bp in length
* 68439
* 68538: gap of unknown length
* 72729: contig of 4191 bp in length
* 68539

```
* 72730 72829: gap of unknown length
* 72830 76599: contig of 3770 bp in length
* 76600 76699: gap of unknown length
* 76700 80689: contig of 3990 bp in length
* 80690 80789: gap of unknown length
* 80790 84157: contig of 3368 bp in length
* 84158 84257: gap of unknown length
* 84258 86897: contig of 2640 bp in length
* 86898 86997: gap of unknown length
* 86998 90634: contig of 3637 bp in length
* 90635 90734: gap of unknown length
* 90735 93888: contig of 3154 bp in length
* 93889 93988: gap of unknown length
* 93989 96425: contig of 2437 bp in length
* 96426 96525: gap of unknown length
* 96526 99226: contig of 2701 bp in length
* 99227 99326: gap of unknown length
* 99327 101527: contig of 2201 bp in length
* 101528 101627: gap of unknown length
* 101628 103979: contig of 2352 bp in length
* 103980 104079: gap of unknown length
* 104080 106536: contig of 2557 bp in length
* 10637 106736: gap of unknown length
* 106737 108746: contig of 2010 bp in length
* 108747 108846: gap of unknown length
* 108847 111300: contig of 2454 bp in length
* 111301 111400: gap of unknown length
* 111401 113117: contig of 1717 bp in length
* 113118 113217: gap of unknown length
* 113218 115167: contig of 1950 bp in length
* 115168 115267: gap of unknown length
* 115268 117132: contig of 1865 bp in length
* 117133 117232: gap of unknown length
* 117233 119189: contig of 1957 bp in length
* 119190 119289: gap of unknown length
* 119290 121294: contig of 2005 bp in length
* 121295 121394: gap of unknown length
* 121395 124027: contig of 2633 bp in length
* 124028 124127: gap of unknown length
* 124128 125297: contig of 1170 bp in length
* 125298 125397: gap of unknown length
* 125398 127248: contig of 1851 bp in length
* 127249 127348: gap of unknown length
* 127349 128892: contig of 1544 bp in length
* 128893 128992: gap of unknown length
* 128993 130141: contig of 1149 bp in length
* 130142 130241: gap of unknown length
* 130242 131814: contig of 1573 bp in length
* 131815 131914: gap of unknown length
* 131915 133894: contig of 1980 bp in length
* 133895 133994: gap of unknown length
* 133995 135143: contig of 1149 bp in length
* 135144 135243: gap of unknown length
* 135244 136514: contig of 1271 bp in length
* 136515 136614: gap of unknown length
* 136615 137962: contig of 1348 bp in length
* 137963 138062: gap of unknown length
* 138063 139600: contig of 1538 bp in length
* 139601 139700: gap of unknown length
* 139701 141172: contig of 1472 bp in length
* 141173 141272: gap of unknown length
* 141273 143083: contig of 1811 bp in length
* 143084 143183: gap of unknown length
* 143184 144605: contig of 1422 bp in length
* 144606 144705: gap of unknown length
* 144706 146421: contig of 1716 bp in length
* 146422 146521: gap of unknown length
* 146522 147819: contig of 1298 bp in length
* 147820 147919: gap of unknown length
* 147920 149416: contig of 1497 bp in length
* 149417 149516: gap of unknown length
* 149517 150980: contig of 1464 bp in length
* 150981 151080: gap of unknown length
```

```
* 151081 152342: contig of 1262 bp in length
* 152343 152442: gap of unknown length
* 152443 153588: contig of 1146 bp in length
* 153589 153688: gap of unknown length
* 153689 155140: contig of 1452 bp in length
* 155141 155240: gap of unknown length
* 155241 156686: contig of 1446 bp in length
* 156687 156786: gap of unknown length
* 156787 158023: contig of 1237 bp in length.

FEATURES
  source          Location/Qualifiers
 1..158023
/organism="Rattus norvegicus"
```

```
Query Match          75.0%; Score 18; DB 2; Length 158023;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      7 ctcacactgtttgttag 24
          |||
Db 36687 CTCACACTGTTTGTCTAG 36670
```

Search completed: August 14, 2002, 21:47:45
Job time: 13483 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 18:06:07 ; Search time 203.42 Seconds
(without alignments)
28,980 Million cell updates/sec

Title: US-09-707-919-1

Perfect score: 24

Sequence: 1 ctccgcctcagactgtttgtag 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	355	4 US-09-043-303-1	Sequence 1, Appl1
2	24	100.0	623	4 US-09-043-303-5	Sequence 5, Appl1
3	24	100.0	4481	4 US-09-041-886-18	Sequence 18, Appl1
4	16	66.7	858	4 US-09-333-521-2	Sequence 2, Appl1
5	15.8	65.8	33	1 US-08-499-048-7	Sequence 7, Appl1
6	15.4	64.2	7032	2 US-08-149-097D-24	Sequence 24, Appl1
7	15.4	64.2	7032	3 US-08-949-386-24	Sequence 24, Appl1
8	15.4	64.2	7032	3 US-08-450-562-24	Sequence 24, Appl1
9	15.4	64.2	7032	4 US-08-984-709A-24	Sequence 24, Appl1
10	15.4	64.2	7089	3 US-08-449-386-25	Sequence 25, Appl1
11	15.4	64.2	7089	3 US-08-450-562-25	Sequence 25, Appl1
12	15.4	64.2	7089	4 US-08-984-709A-25	Sequence 25, Appl1
13	15.2	63.3	1247	1 US-08-773-870-2	Sequence 2, Appl1
14	15.2	63.3	6270	1 US-08-418-893D-25	Sequence 25, Appl1
15	15.2	63.3	6790	1 US-08-418-893D-22	Sequence 22, Appl1
16	15.2	63.3	6946	4 US-09-316-080-1	Sequence 1, Appl1
17	15	62.5	358	1 US-07-841-652-2	Sequence 2, Appl1
18	15	62.5	1302	1 US-08-425-299A-1	Sequence 1, Appl1
19	15	62.5	1452	1 US-08-187-785-2	Sequence 2, Appl1
20	15	62.5	1501	3 US-09-046-736-1	Sequence 1, Appl1
21	15	62.5	1502	3 US-09-046-736-3	Sequence 3, Appl1
22	15	62.5	7620	1 US-07-767-135-1	Sequence 1, Appl1
23	15	62.5	7620	1 US-07-841-652-1	Sequence 1, Appl1
24	15	62.5	9785	4 US-09-479-128-1	Sequence 1, Appl1
25	14.8	61.7	1621	3 US-08-933-750C-96	Sequence 96, Appl1
26	14.8	61.7	1621	3 US-09-234-613-96	Sequence 96, Appl1
27	14.8	61.7	2688	1 US-08-088-633-3	Sequence 3, Appl1

c 28	14.8	61.7	2688	1 US-08-245-756-3	Sequence 3, Appl1
c 29	14.8	61.7	2688	1 US-08-441-750-3	Sequence 3, Appl1
c 30	14.8	61.7	2688	1 US-08-441-751-3	Sequence 3, Appl1
c 31	14.8	61.7	2688	5 PCR-US92-02521-3	Sequence 3, Appl1
c 32	14.8	61.7	2840	5 PCR-US93-06251-64	Sequence 64, Appl1
c 33	14.8	61.7	2875	5 PCR-US93-06251-63	Sequence 63, Appl1
c 34	14.6	60.8	585	3 US-08-938-546B-5	Sequence 5, Appl1
c 35	14.6	60.8	585	4 US-08-938-093A-5	Sequence 5, Appl1
c 36	14.6	60.8	1317	3 US-08-886-886-1	Sequence 1, Appl1
c 37	14.6	60.8	8224	6 5180808-1	Patent No. 5180808
c 38	14.6	60.8	36741	4 US-09-301-665-3	Sequence 3, Appl1
c 39	14.4	60.0	477	2 US-08-653-402B-1	Sequence 1, Appl1
c 40	14.4	60.0	713	2 US-08-365-486A-8	Sequence 8, Appl1
c 41	14.4	60.0	713	4 US-08-880-342-8	Sequence 8, Appl1
c 42	14.4	60.0	1245	4 US-09-282-305-15	Sequence 15, Appl1
c 43	14.4	60.0	1308	1 US-08-395-742-2	Sequence 2, Appl1
c 44	14.4	60.0	2712	1 US-08-346-455B-37	Sequence 37, Appl1
c 45	14.4	60.0	2712	3 US-08-977-221-37	Sequence 37, Appl1

ALIGNMENTS

```

RESULT 1
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
; US-09-043-303-1

```

```

Query Match          100.0%; Score 24; DB 4; Length 355;
Best Local Similarity 100.0%; Pred No. 0.0031;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctccgcctcagactgtttgtag 24
|||||
Db 73 ctccgcctcagactgtttgtag 96

```

```

RESULT 2
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17

```


RESULT 6
US-08-149-097D-24
Sequence 24, Application US/08149097D
Patent No. 5874236
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,097D
FILING DATE: 05-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/482,384
FILING DATE: 20-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/176,899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-55038
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 7032 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 169..6921
OTHER INFORMATION: /product="Alphale-1 subunit of
OTHER INFORMATION: human calcium channel"
US-08-149-097D-24
Query Match 64.2%; Score 15.4; DB 2; Length 7032;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 cgcctcagactgttttg 20
Db 6699 CGCCTCAGACTGTGTG 6715
RESULT 7
US-08-949-386-24
Sequence 24, Application US/08949386
Patent No. 6090623
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Gillespie, Alison
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/949,386
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/290,012
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 519808
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 7032 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6921

OTHER INFORMATION: /standard_name="Alpha-1E-1"
US-08-949-386-24

Query Match 64.2% Score 15.4; DB 3; Length 7032;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 cgcctcagactgtttg 20
|||||
DB 6699 CGCCTCAGACTGTGTG 6715

RESULT 8
US-08-450-562-24
Sequence 24, Application US/08450562
Patent No. 6096514
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Gillespie, Allison
APPLICANT: Feldman, Daniel
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,562
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/404,950
FILING DATE: 13-MAR-1995
APPLICATION NUMBER: 08/336,257
FILING DATE: 7-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,363
FILING DATE: 23-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,012
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/223,305
FILING DATE: 4-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/193,078
FILING DATE: 07-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,231
FILING DATE: 13-JULY-1992
COMPUTER READABLE FORM:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/603,751
FILING DATE: 08-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/482,384
FILING DATE: 02-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/176,899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-519812
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 7032 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6921
OTHER INFORMATION: /standard_name="Alpha-1E-1"
US-08-450-562-24

Query Match 64.2% Score 15.4; DB 3; Length 7032;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 cgcctcagactgtttg 20
|||||
DB 6699 CGCCTCAGACTGTGTG 6715

RESULT 9
US-08-984-709A-24
Sequence 24, Application US/08984709A
Patent No. 6320032
GENERAL INFORMATION:
APPLICANT: Williams, Mark E.
APPLICANT: Stauderman, Kenneth A.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehtman White & Mcauliffe
STREET: 4250 Executive Square, Suite 700
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,709A
FILING DATE: 02-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 450-8400
TELEFAX: (619) 587-5360
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 7032 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6921
OTHER INFORMATION: /standard_name="Alpha-1E-1"
US-08-984-709A-24

Query Match 64.2%; Score 15.4; DB 4; Length 7032;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cgcctcagactgtttg 20
|||||

DB 6699 CGCCTCAGACTGTGTG 6715

RESULT 10
US-08-949-386-25
Sequence 25, Application US/08949386
Patent No. 6090623
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Gillespie, Allison
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/949,386
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/290,012
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 519808
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 7089 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6978
OTHER INFORMATION: /standard_name="Alpha-1E-3"
US-08-949-386-25

Query Match 64.2%; Score 15.4; DB 3; Length 7089;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 cgcctcagactgtttg 20
|||||

DB 6756 CGCCTCAGACTGTGTG 6772

RESULT 11
US-08-450-562-25
Sequence 25, Application US/08450562
Patent No. 6096514
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Gillespie, Allison
APPLICANT: Feldman, Daniel
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,562
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/404,950
FILING DATE: 13-MAR-1995
APPLICATION NUMBER: 08/336,257
FILING DATE: 7-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,363

PRIOR FILING DATE: 23-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,012
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/223,305
FILING DATE: 4-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/193,078
FILING DATE: 07-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,231
FILING DATE: 13-JULY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/603,751
FILING DATE: 08-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/482,384
FILING DATE: 02-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/176,899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-519812
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 7089 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6978
OTHER INFORMATION: /standard_name="Alpha-1E-3"
US-08-450-562-25

Query Match 64.2% Score 15.4; DB 3; Length 7089;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 cgccctcagactgttttg 20
|||||
Db 6756 CGCCTCAGACTGTGTG 6772

RESULT 12
US-08-984-709A-25
Sequence 25, Application US/08984709A
Patent No. 6320032
GENERAL INFORMATION:
APPLICANT: Williams, Mark E.
APPLICANT: Stauderman, Kenneth A.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehrman White & McCauliffe
STREET: 4250 Executive Square, Suite 700
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,709A
FILING DATE: 02-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 450-8400
TELEFAX: (619) 587-5360
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 7089 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 166..6978
OTHER INFORMATION: /standard_name="Alpha-1E-3"
US-08-984-709A-25

Query Match 64.2% Score 15.4; DB 4; Length 7089;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 cgccctcagactgttttg 20
|||||
Db 6756 CGCCTCAGACTGTGTG 6772

RESULT 13
US-08-773-870-2
Sequence 2, Application US/08773870
Patent No. 5912143
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN MAGE-LIKE PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773,870
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0179 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1247 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
US-08-773-870-2

Query Match      63.3%; Score 15.2; DB 2; Length 1247;
Best Local Similarity 85.0%; Pred. No. 98;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      5 gccctcagactgtttgtag 24
        ||||||| ||||| ||| ||
Db      179 gccctcagatgttgcgag 198

RESULT 14
US-08-418-893D-25/c
Sequence 25, Application US/08418893D
Patent No. 5559220
GENERAL INFORMATION:
APPLICANT: ROESSLER, PAUL G
APPLICANT: OHLROGE, JOHN B
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
STREET: 1617 Cole Blvd.
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,893D
FILING DATE: April 7, 1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/104,938
FILING DATE: September 14, 1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: O'CONNOR, EDNA
```

```
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-231-1000
TELEFAX: 303-231-1098
TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 6270 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-418-893D-25
```

```
Query Match      63.3%; Score 15.2; DB 1; Length 6270;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      5 gccctcagactgtttgtag 24
        ||||||| ||||| ||| ||
Db      3836 gcattcagactgttcgag 3817
```

```
RESULT 15
US-08-418-893D-22/c
Sequence 22, Application US/08418893D
Patent No. 5559220
GENERAL INFORMATION:
APPLICANT: ROESSLER, PAUL G
APPLICANT: OHLROGE, JOHN B
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
STREET: 1617 Cole Blvd.
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,893D
FILING DATE: April 7, 1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/104,938
FILING DATE: September 14, 1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: O'CONNOR, EDNA
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-231-1000
TELEFAX: 303-231-1098
TELEX:
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 6790 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
```

! ANTI-SENSE: NO
US-08-418-893D-22

Query Match 63.3%; Score 15.2; DB 1; Length 6790;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gccctcagactgttttgtag 24
|||
DB 4356 GCATCAGACTGTTCCTGGGAG 4337

Search completed: August 14, 2002, 21:50:47
Job time: 13480 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 18:00:52 ; Search time 7749.14 Seconds
(without alignments)
41.802 Million cell updates/sec

Title: US-09-707-919-1

Perfect score: 24

Sequence: 1 cccgcctcagactgtttgtag 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estropl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrti:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	24	100.0	482 9	AL039573 DKFZp434D
2	24	100.0	500 10	BI547486 603191091
3	23	95.8	126 10	F14808 SSC20D02 Po
4	19.4	80.8	885 3	BF234301 602026211
5	19.2	80.0	646 9	AU081685 AU081685
6	19.2	80.0	998 12	CNS04F01 Tetradon
7	19.2	79.2	1100 10	BM455214 AGENCOURT
8	18.4	76.7	503 9	AI484086 EST249957
9	18.4	76.7	612 9	AI776858 EST25958
10	18.4	76.7	659 10	BI923981 EST543870
11	18.4	76.7	660 10	BI932971 EST52860
12	18.4	76.7	738 10	BI923984 EST543873
13	18.4	76.7	760 10	BI925151 EST545040
14	18.4	76.7	765 10	BI926315 EST546204
15	18.4	76.7	1004 12	CNS055XW Tetradon
16	18.2	75.8	385 10	BM308548 sak47e04
17	17.8	74.2	356 12	AZ307058 IM0008H19

18	17.8	74.2	418 12	A0882650 HS_5431-B
19	17.6	73.3	133 12	BI199208 TC3-24G6
20	17.6	73.3	601 10	BF611779 de88f12.y
21	17.6	73.3	654 9	BB087977 BB087977
22	17.6	73.3	800 12	BH584322 BOCJX69TF
23	17.6	73.3	1018 10	BF681972 602116941
24	17.4	72.5	489 10	BG351014 099H07 Ma
25	17.4	72.5	577 9	AJ397058
26	17.4	72.5	588 12	AQ305461 HS_2046-A
27	17.4	72.5	617 10	BI513100 BRL60011A
28	17.4	72.5	1020 12	CNS033PR Tetradon
29	17.2	71.7	302 9	BB375200 BB375200
30	17.2	71.7	430 10	BE342148 EST394990
31	17.2	71.7	544 10	BC098471 EST462990
32	17.2	71.7	547 9	AM033668 EST277239
33	17.2	71.7	582 12	AZ830700 2M0110A08
34	17.2	71.7	598 10	BJ043218 BJ043218
35	17.2	71.7	601 12	AZ822647 2M0096N01
36	17.2	71.7	665 9	AI729728 BMLG1140
37	17.2	71.7	764 10	BF683315 602139458
38	17.2	71.7	938 10	BF103355 601646787
39	17.2	71.7	942 10	BC291680 602385831
40	17.2	71.7	1051 12	CNS05HV7 Tetradon
41	17.2	71.7	1111 12	CNS04C03 Tetradon
42	17.2	71.7	1291 12	AG055213 Pan trogl
43	17.2	71.7	1584 10	BE965666 601659719
44	17.2	71.7	1789 10	BE966241 601660177
45	17.2	70.8	470 12	AQ798476 HS_3212-B

ALIGNMENTS

RESULT 1
LOCUS AL039573 482 bp mRNA linear EST 29-FEB-2000
DEFINITION DKFZp434D1311.t1 434 (synonym: htes3) Homo sapiens CDNA clone
ACCESSION AL039573
VERSION AL039573.1 GI:5408612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Homo sapiens Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 482)
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
,S. (Duesterhoeft, et al.)
EST (Duesterhoeft, et al.)
Unpublished (1999)
Contact: Duesterhoeft A
MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
sequenced by Olegen (Hilden/Germany) within the CDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp434D1311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubneweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.
location/Qualifiers
1. 482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434D1311"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

FEATURES

source

BASE COUNT

49 a 218 c 145 g 70 t

ORIGIN

Query Match 100.0%; Score 24; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24
|||||
Db 22 CTCGCCCTCAGACTGTTTGGTAG 45

RESULT 2
BIS47486 500 bp mRNA linear EST 05-SEP-2001
LOCUS 603191091P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262335 5',
DEFINITION mRNA sequence.
ACCESSION BIS47486
VERSION BIS47486.1 GI:15434798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs.rem@nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM1161 row: e column: 24
High quality sequence stop: 485.

FEATURES
source
Location/Qualifiers
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262335"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgcag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 57 a 222 c 150 g 71 t

ORIGIN

Query Match 100.0%; Score 24; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24
|||||
Db 25 CTCGCCCTCAGACTGTTTGGTAG 48

RESULT 3
F14808 126 bp mRNA linear EST 09-SEP-1996
LOCUS

DEFINITION SSC20D02 Porcine small intestine cDNA library Sus scrofa cDNA clone
C20D02, mRNA sequence.
ACCESSION F14808
VERSION F14808.1 GI:971822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 126)
AUTHORS Wintero, A.K., Fredholm, M. and Davies, W.
TITLE Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones
JOURNAL Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE 96327607
COMMENT Contact: A.K. Wintero
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowvej 13, 1870 Frederiksberg C, Denmark.

FEATURES
source
Location/Qualifiers
1..126
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="C20D02"
/clone_lib="porcine small intestine cDNA library"
/note="directionally cloned cDNA in XLI-blue MRF"

BASE COUNT 9 a 54 c 37 g 24 t

ORIGIN

Query Match 95.8%; Score 23; DB 10; Length 126;
Best Local Similarity 95.8%; Pred. No. 1.2;
Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ctccgcctcagactgtttgtag 24
|||||
Db 6 CTCGCCCTCAGACTGTTTGGTAG 29

RESULT 4
BF234301 885 bp mRNA linear EST 14-NOV-2000
LOCUS 60202611P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4161411 5',
DEFINITION mRNA sequence.
ACCESSION BF234301
VERSION BF234301.1 GI:1145643
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs.rem@nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9442 row: f column: 04
High quality sequence stop: 755.

FEATURES
source
Location/Qualifiers
1..885
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4161411"
/clone_lib="NCI_CGAP_L19"

/lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: PCMV-SPORT6; Site_1: Not;
 Site_2: Salt: Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: This is a NCI-CGAP Library."
 BASE COUNT 212 a 264 c 223 g 186 t
 ORIGIN

Query Match 80.8%; Score 19.4; DB 10; Length 885;
 Best Local Similarity 95.2%; Pred. No. 1e+02; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctccgcctcagactgtttgtag 21
 |||||
 Db 630 CTCGCCCTCAGACTGTATTGG 650

RESULT 5
 AU081685 646 bp mRNA linear EST 08-JUN-2001
 LOCUS AU081685 Marchantia polymorpha strain E sexual organ, antheridium,
 DEFINITION immature Marchantia polymorpha cDNA clone M010042, mRNA sequence.
 ACCESSION AU081685
 VERSION AU081685.1 GI:14329436
 KEYWORDS EST.
 SOURCE liverwort.
 ORGANISM Marchantia polymorpha
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta;
 Marchantiophyta; Marchantiopsida; Marchantiidae; Marchantiales;
 Marchantiaceae; Marchantia.
 TITLE 1 (bases 1 to 646)
 Nishiyama, R., Yamato, K.T., Miura, K., Sakaida, M., Okada, S., Kono, K.,
 AUTHORs Takahama, M., Sone, T., Takenaka, M., Fukuzawa, H. and Ohyama, K.
 Comparison of expressed sequence tags from male and female sexual
 organs of Marchantia polymorpha
 JOURNAL DNA Res 7, 165-174 (2000)
 MEDLINE 20363092
 COMMENT Contact: Katsuyuki T Yamato
 Graduate School of Biostudies, Division of Integrated Life Science
 Kyoto University
 Sakyo-Ku, Kitashirakawaoiwake-cho, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-6453
 Fax: 81-75-753-6127
 Email: kyamato@lif.kyoto-u.ac.jp
 other clones with same sequence: M010046, M01S085.
 Location/Qualifiers

FEATURES
 source
 1..646
 /organism="Marchantia polymorpha"
 /strain="E"
 /db_xref="taxon:3197"
 /clone="M010042"
 /clone_11b="Marchantia polymorpha strain E sexual organ,
 antheridium, immature"
 /sex="male"
 /tissue.type="sexual organ, antheridium, immature"
 BASE COUNT 153 a 140 c 182 g 171 t
 ORIGIN

Query Match 80.0%; Score 19.2; DB 9; Length 646;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ctccgcctcagactgtttgtag 24
 |||||
 Db 417 CTCGCCCTCAGACTGTATTGG 440

RESULT 6
 CNS04RDI 998 bp DNA linear GSS 24-MAY-2000
 LOCUS CNS04RDI/c
 DEFINITION Tetraodon nigroviridis genome survey sequence 77 end of clone

003C22 of library H from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL303759
 VERSION AL303759.1 GI:8188364
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 998)
 Roest-Crolius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 AUTHORs Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis

TITLE Unpublished
 JOURNAL 2 (bases 1 to 998)
 Roest-Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 AUTHORs Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

REFERENCE 3 (bases 1 to 998)
 Genoscope.
 Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/tetraodon.

FEATURES
 source
 1..998
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="003C22"
 /clone_11b="H"
 /note="Genoscope sequence ID : C08H003B11XN1-end : 77"
 BASE COUNT 295 a 225 c 189 g 288 t 1 others
 ORIGIN

Query Match 80.0%; Score 19.2; DB 12; Length 998;
 Best Local Similarity 87.5%; Pred. No. 1.3e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ctccgcctcagactgtttgtag 24
 |||||
 Db 653 CTCGACATCAGACTGTATTGGTAG 630

RESULT 7
 BM455214 1100 bp mRNA linear EST 05-FEB-2002
 LOCUS BM455214
 DEFINITION AGENCOURT 6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
 5' mRNA sequence.
 ACCESSION BM455214
 VERSION BM455214.1 GI:18504254
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1100)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORs National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE JOURNAL
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Lou Staudt
 cDNA library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
Plate: L1AM12134 row: k column: 12
High quality sequence stop: 623.

FEATURES

SOURCE

1. 1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DHI0B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 240 a 329 c 306 g 219 t 6 others
ORIGIN

Query Match 79.2%; Score 19; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 cctcagactgtttgttag 24
|||||
Db 1 CCTCAGACTGTTTGTGTAG 19

RESULT 8

LOCUS A1484086 503 bp mRNA linear EST 18-MAY-2001
DEFINITION EST249957 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLED2513, mRNA sequence.

ACCESSION A1484086
VERSION A1484086.1 GI:438010
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.
1 (bases 1 to 503)

REFERENCE 1 (bases 1 to 503)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)

TITLE

JOURNAL
COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES

SOURCE

1. 503
/organism="Lycopersicon esculentum"
/cultivar="T496"
/db_xref="taxon:4081"
/clone="CLED2513"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="X11-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,

BASE COUNT 130 a 117 c 109 g 147 t
ORIGIN

Query Match 76.7%; Score 18.4; DB 9; Length 503;
Best Local Similarity 95.0%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 gccctcagactgtttgttag 24
|||||
Db 6 GCCCTCAGACTGTTTGTGTAG 25

RESULT 9

LOCUS A1776858 612 bp mRNA linear EST 18-MAY-2001
DEFINITION EST257958 tomato resistant, Cornell Lycopersicon esculentum cDNA
clone CLERR20C21, mRNA sequence.

ACCESSION A1776858
VERSION A1776858.1 GI:5274899
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE

1 (bases 1 to 612)
D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
Giovannoni,J.J. and Martin,G.B.
Generation of ESTs from Pseudomonas resistant tomato
Unpublished (1999)

TITLE

JOURNAL
COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES

SOURCE

1. 612
/organism="Lycopersicon esculentum"
/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="CLERR20C21"
/clone_lib="tomato resistant, Cornell"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library
directionally cloned cDNAs inserted into pBluescript SK(-)
at 5' end with EcoRI and 3' end with XhoI site."
BASE COUNT 176 a 113 c 146 g 177 t
ORIGIN

Query Match 76.7%; Score 18.4; DB 9; Length 612;
Best Local Similarity 95.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 gccctcagactgtttgttag 24
|||||
Db 547 GCCCTCAGACTGTTTGTGTAG 566

RESULT 10
LOCUS B1923981 659 bp mRNA linear EST 18-OCT-2001
DEFINITION EST543870 tomato flower, buds 0-3 mm Lycopersicon esculentum cDNA
clone CTOA21N10 5' end, mRNA sequence.

```

ACCESSION   B1923981
VERSION      B1923981.1  GI:16225917
KEYWORDS
SOURCE       tomato.
ORGANISM     Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE    1 (bases 1 to 659)
AUTHORS      van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
              Uterback,T., Van Aken,S., Romling,C.M., Nierman,W., Fraser,C.M.,
              Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
              Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)
              Unpublished (2001)
COMMENT      Contact: CUGI
              Clemson University Genomics Institute
              100 Jordan Hall, Clemson, SC 29634, USA
              Email: http://www.genome.clemson.edu/orders/index.html
              This clone is available through the Clemson University Genomics
              Institute
FEATURES
  source      Location/Qualifiers
              1..659
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CT0A21N10"
                /clone_1lb="tomato flower, buds 0-3 mm"
                /tissue_type="flower"
                /dev_stage="0-3mm buds"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University; sequencing: The
                Institute for Genomic Research; flower buds and flowers
                were taken from greenhouse plants (4-8 wks old, TA496).
                They were immediately frozen in liquid nitrogen and then
                size-separated while remaining frozen."
BASE COUNT   195 a 121 c 147 g 196 t
ORIGIN
Query Match      76.7%; Score 18.4; DB 10; Length 659;
Best Local Similarity 95.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 gccctcagactgtttgtgtag 24
||||||| |||||||||
Db 616 gccctcagaaatgtttgtgtac 635

RESULT 11
B1923971 LOCUS      B1923971 660 bp mRNA linear EST 18-OCT-2001
DEFINITION      EST552860 tomato flower, 8 mm to preanthesis buds Lycopersicon
                  esculentum cDNA clone cT0C24F10 5' end, mRNA sequence.
ACCESSION      B1923971
VERSION        B1923971.1  GI:16247443
KEYWORDS
SOURCE         tomato.
ORGANISM       Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE      1 (bases 1 to 660)
AUTHORS        van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
                Uterback,T., Van Aken,S., Romling,C.M., Nierman,W., Fraser,C.M.,
                Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
                Generation of ESTs from tomato flower tissue, buds 8 mm -
                preanthesis
                Unpublished (2001)
COMMENT        Contact: CUGI
                Clemson University Genomics Institute
                100 Jordan Hall, Clemson, SC 29634, USA
                Email: http://www.genome.clemson.edu/orders/index.html
                This clone is available through the Clemson University Genomics
                Institute
FEATURES
  source      Location/Qualifiers
              1..738
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CT0A21N22"
                /clone_1lb="tomato flower, buds 0-3 mm"
                /tissue_type="flower"
                /dev_stage="0-3mm buds"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University; sequencing: The
                Institute for Genomic Research; flower buds and flowers
                were taken from greenhouse plants (4-8 wks old, TA496).
                They were immediately frozen in liquid nitrogen and then
                size-separated while remaining frozen."
BASE COUNT   190 a 126 c 145 g 199 t
ORIGIN
Query Match      76.7%; Score 18.4; DB 10; Length 660;
Best Local Similarity 95.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 gccctcagactgtttgtgtag 24
||||||| |||||||||
Db 635 gccctcagaaatgtttgtgtac 654

RESULT 12
B1923984 LOCUS      B1923984 738 bp mRNA linear EST 18-OCT-2001
DEFINITION      EST543873 tomato flower, buds 0-3 mm Lycopersicon esculentum cDNA
                  clone cT0A21N22 5' end, mRNA sequence.
ACCESSION      B1923984
VERSION        B1923984.1  GI:16225997
KEYWORDS
SOURCE         tomato.
ORGANISM       Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE      1 (bases 1 to 738)
AUTHORS        van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
                Uterback,T., Van Aken,S., Romling,C.M., Nierman,W., Fraser,C.M.,
                Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
                Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)
                Unpublished (2001)
COMMENT        Contact: CUGI
                Clemson University Genomics Institute
                100 Jordan Hall, Clemson, SC 29634, USA
                Email: http://www.genome.clemson.edu/orders/index.html
                This clone is available through the Clemson University Genomics
                Institute
FEATURES
  source      Location/Qualifiers
              1..738
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CT0A21N22"
                /clone_1lb="tomato flower, buds 0-3 mm"
                /tissue_type="flower"
                /dev_stage="0-3mm buds"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University; sequencing: The
                Institute for Genomic Research; flower buds and flowers
                were taken from greenhouse plants (4-8 wks old, TA496).
                They were immediately frozen in liquid nitrogen and then
                size-separated while remaining frozen."
BASE COUNT   190 a 126 c 145 g 199 t
ORIGIN

```

```

ACCESSION   B1923981
VERSION      B1923981.1  GI:16225917
KEYWORDS
SOURCE       tomato.
ORGANISM     Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE    1 (bases 1 to 659)
AUTHORS      van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
              Uterback,T., Van Aken,S., Romling,C.M., Nierman,W., Fraser,C.M.,
              Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
              Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)
              Unpublished (2001)
COMMENT      Contact: CUGI
              Clemson University Genomics Institute
              100 Jordan Hall, Clemson, SC 29634, USA
              Email: http://www.genome.clemson.edu/orders/index.html
              This clone is available through the Clemson University Genomics
              Institute
FEATURES
  source      Location/Qualifiers
              1..659
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CT0A21N10"
                /clone_1lb="tomato flower, buds 0-3 mm"
                /tissue_type="flower"
                /dev_stage="0-3mm buds"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University; sequencing: The
                Institute for Genomic Research; flower buds and flowers
                were taken from greenhouse plants (4-8 wks old, TA496).
                They were immediately frozen in liquid nitrogen and then
                size-separated while remaining frozen."
BASE COUNT   190 a 126 c 145 g 199 t
ORIGIN
Query Match      76.7%; Score 18.4; DB 10; Length 660;
Best Local Similarity 95.0%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 gccctcagactgtttgtgtag 24
||||||| |||||||||
Db 635 gccctcagaaatgtttgtgtac 654

RESULT 12
B1923984 LOCUS      B1923984 738 bp mRNA linear EST 18-OCT-2001
DEFINITION      EST543873 tomato flower, buds 0-3 mm Lycopersicon esculentum cDNA
                  clone cT0A21N22 5' end, mRNA sequence.
ACCESSION      B1923984
VERSION        B1923984.1  GI:16225997
KEYWORDS
SOURCE         tomato.
ORGANISM       Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE      1 (bases 1 to 738)
AUTHORS        van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
                Uterback,T., Van Aken,S., Romling,C.M., Nierman,W., Fraser,C.M.,
                Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
                Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)
                Unpublished (2001)
COMMENT        Contact: CUGI
                Clemson University Genomics Institute
                100 Jordan Hall, Clemson, SC 29634, USA
                Email: http://www.genome.clemson.edu/orders/index.html
                This clone is available through the Clemson University Genomics
                Institute
FEATURES
  source      Location/Qualifiers
              1..738
                /organism="Lycopersicon esculentum"
                /cultivar="TA496"
                /db_xref="taxon:4081"
                /clone="CT0A21N22"
                /clone_1lb="tomato flower, buds 0-3 mm"
                /tissue_type="flower"
                /dev_stage="0-3mm buds"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI; supplier: Cornell University; sequencing: The
                Institute for Genomic Research; flower buds and flowers
                were taken from greenhouse plants (4-8 wks old, TA496).
                They were immediately frozen in liquid nitrogen and then
                size-separated while remaining frozen."
BASE COUNT   190 a 126 c 145 g 199 t
ORIGIN

```

XhoI: supplier: Cornell University; sequencing: The Institute for Genomic Research; flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT 214 a 136 c 170 g 218 t

Query Match 76.7%; Score 18.4; DB 10; Length 738;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 gccctcagactgtttgttag 24
||||||| |||||||||

Db 616 GCCTCAGAAATGTTTGTAG 635

RESULT 13
BI925151

LOCUS EST545040 tomato flower, buds 0-3 mm Lycopersicon esculentum cDNA
DEFINITION clone cTOA25L23 5' end, mRNA sequence.

ACCESSION BI925151 GI:16230129

VERSION EST.

KEYWORDS tomato.

SOURCE Lycopersicon esculentum

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 760)

REFERENCE van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
AUTHORS Uiterback,T., Van Aken,S., Roming,C.M., Niernman,W., Fraser,C.M.,
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)

Unpublished (2001)

CONTACT: CUGI

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

FEATURES Location/Qualifiers

1..760
/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cTOA25L23"

/clone.lib="tomato flower, buds 0-3 mm"

/tissue_type="flower"

/dev_stage="0-3mm buds"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Cornell University; sequencing: The

Institute for Genomic Research; flower buds and flowers

were taken from greenhouse plants (4-8 wks old, TA496).

They were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 215 a 146 c 171 g 228 t

ORIGIN

Query Match 76.7%; Score 18.4; DB 10; Length 760;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 gccctcagactgtttgttag 24
||||||| |||||||||

Db 634 GCCTCAGAAATGTTTGTAG 653

RESULT 14
BI926315 765 bp mRNA linear EST 18-OCT-2001
LOCUS EST546204 tomato flower, buds 0-3 mm Lycopersicon esculentum cDNA
DEFINITION clone cTOA29E20 5' end, mRNA sequence.

ACCESSION BI926315 GI:16234506

VERSION EST.

KEYWORDS tomato.

SOURCE Lycopersicon esculentum

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 765)

REFERENCE van der Hoeven,R.S., Bezzerides,J.L., Karamycheva,S.A., Tsai,J.,
AUTHORS Uiterback,T., Van Aken,S., Roming,C.M., Niernman,W., Fraser,C.M.,
Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.

Generation of ESTs from tomato flower tissue, 0-3 mm buds (2001)

Unpublished (2001)

CONTACT: CUGI

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

FEATURES Location/Qualifiers

1..765
/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cTOA29E20"

/clone.lib="tomato flower, buds 0-3 mm"

/tissue_type="flower"

/dev_stage="0-3mm buds"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Cornell University; sequencing: The

Institute for Genomic Research; flower buds and flowers

were taken from greenhouse plants (4-8 wks old, TA496).

They were immediately frozen in liquid nitrogen and then

size-separated while remaining frozen."

BASE COUNT 216 a 148 c 172 g 229 t

ORIGIN

Query Match 76.7%; Score 18.4; DB 10; Length 765;
Best Local Similarity 95.0%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 gccctcagactgtttgttag 24
||||||| |||||||||

Db 634 GCCTCAGAAATGTTTGTAG 653

RESULT 15
CNS055XW

LOCUS CNS055XW 1004 bp DNA linear GSS 26-JUL-2000

DEFINITION Tetradodon nigroviridis genome survey sequence SP6 end of clone

001K06 of library B from Tetradodon nigroviridis, genomic survey

sequence.

ACCESSION AL322637

VERSION AL322637.1 GI:9555521

KEYWORDS GSS: genome survey sequence.

SOURCE Tetradodon nigroviridis.

ORGANISM Tetradodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetradodon.

1 (bases 1 to 1004)

REFERENCE

AUTHORS Roest Croillius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.
 TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
 MEDLINE 20296633
 REFERENCE 2 (bases 1 to 1004)
 AUTHORS Croillius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis
 JOURNAL Genome Res. 10 (7), 939-949 (2000)
 MEDLINE 20359837
 REFERENCE 3 (bases 1 to 1004)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.
 FEATURES
 source Location/Qualifiers
 1..1004
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="001K05"
 /clone_lib="B"
 /note="Genoscope sequence ID : COAB001BF03B1-end : SP6"
 BASE COUNT 248 a 211 c 217 g 285 t 43 others
 ORIGIN
 Query Match 76.7%; Score 18.4; DB 12; Length 1004;
 Best Local Similarity 95.0%; Pred. No. 3.2e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 tcgcgcacagactgttttg 21
 ||| ||||| ||||| ||||| |||||
 Db 610 TCCTCCTCAGACTGTTTTGG 629

Search completed: August 14, 2002, 21:04:00
 Job time: 10988 sec

KEYWORDS	unidentified.
SOURCE	unidentified
ORGANISM	unclassified
REFERENCE	1 (bases 1 to 4200)
AUTHORS	Tora,L., Lutz,Y., Troitier,Y., Mandel and Jean-Louis .
TITLE	METHOD FOR TREATING NEURODEGENERATIVE DISEASES USING A 1C2 ANTIBODY OR A FRAGMENT OR DERIVATIVE THEREOF, AND CORRESPONDING PHARMACEUTICAL COMPOSITIONS
JOURNAL	PATENT: WO 9717445-A 7 15-MAY-1997;
COMMENT	CENTRE NAT RECH SCIENT (FR)
FEATURES	Other publication FR 2741088 19970516. Location/Qualifiers 1..4200 /organism="unidentified" /db_xref="taxon:32644" /clone="DANI"
BASE COUNT	1152 a 1200 c 913 g 935 t
ORIGIN	
Query Match	100.0%; Score 20; DB 6; Length 4200;
Best Local Similarity	100.0%; Pred. No. 1.2e+02;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 gtggccgagacgaggagac 20
Db	433 GTGGCCGAGACGAGAGAC 414
RESULT	6
LOCUS	AR153580/c
DEFINITION	Sequence 18 from patent US 6235872.
ACCESSION	AR153580
VERSION	AR153580.1 GI:15121112
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 4481)
AUTHORS	Bredesen,D.E. and Rabizadeh,S.
TITLE	Pitocaprotic peptides dependence polypeptides and methods of use
JOURNAL	PATENT: US 6235872-A 18 22-MAY-2001;
FEATURES	Location/Qualifiers 1..4481 /organism="unknown"
BASE COUNT	1144 a 1380 c 1014 g 943 t
ORIGIN	
Query Match	100.0%; Score 20; DB 6; Length 4481;
Best Local Similarity	100.0%; Pred. No. 1.2e+02;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 gtggccgagacgaggagac 20
Db	833 GTGGCCGAGACGAGAGAC 814
RESULT	7
LOCUS	HSU70323/c
DEFINITION	Human ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION	U70323
VERSION	U70323.1 GI:1679683
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 4481)
AUTHORS	Pulst,S.-M., Nechiporuk,A., Nechiporuk,T., Glisbert,S., Chen,X.-N.,

TITLE		Lopes-Cendes,I., Pearlman,S., Starkman,S., Orozco-Diaz,G., Lunkes,A., Delong,P., Rouleau,G.A., Auburger,G., Korenberg,J.R., Figueroa,C. and Saba,S.
JOURNAL		Moderate expansion of a normally diallelic trinucleotide repeat in spinocerebellar ataxia type 2
REFERENCE		Nature Genet. 14 (3), 269-276 (1996)
AUTHORS		Puls,T.S.-M.
TITLE		(bases 1 to 4481)
JOURNAL		Submitted (10-SEP-1996) Medicine, Cedars-Sinai, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
FEATURES		Location/Qualifiers 1..4481
Source		/organism-"Homo sapiens" /db_xref-"taxon:9606" /chromosome-"12" /map-"1q24.1" 1..4481 /gene-"SCA2" 163..4101 /gene-"SCA2" /_strandname-"spinocerebellar ataxia type 2" /_codon_start=1 /_product-"ataxin-2" /_protein_id-"AB19200.1" /_db_xref-"GI:1679684" /_translation-"MSRAAAPRSPAVATESRRPFAARMPWRSLQPARNSRGHGCGGAAAPGPSPGGPPSGPSASPCFCSNNGGAFFPGSRRLGLLGCPPLPAFPVYLILASPAGPAPPPTPRPASPIGRASRGVSIGRAPCCPDCEPVYGLTRMSLRKPOOOOONNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSSSAATSVVAATISGGRGLDGRSNLGNLOSTI SEDGITANRWNYHLTSYGSKEEVOYNKYEGVKRTYSPKDVLDAHEKSTESSPKKEETMESILECSDFOVVVFVKMDSYAKRDFTDSIAIKANGEKEDLPWAGELTANEDELAEINDESSNGMDPNDFERNEENTGVSYIDLSLSITVPLEDRNSEEFLLAREANOIAEIEISSAOXKARYALENDRESEEXYTAVONSERGEHSINTRENKIIPGO RNREVISMGSGRROMNPBGOPGSGMSGRSHTSDFNPNHQSQALNVGCYPWPSPGPSRPSTRYSOSCNLSLPPRAATTTPRSSPRSPPSRSPRSBSMGSPAUSPTKMSSSGPMARSKAQHNRHNRYASARGSTSGLTFVSHSPBEALTTPVARTISPFGCTIVSYGEPRLSPKHHRPSRPONSIGNTPSGVLAHPDAGITPEEAAMPRIAPASPTRPASVRAYT PPSKANDRLDOORONSPAGKNENIKPRETSPSPSKENKGISPYVESENKOIDLKRKFKNPFRLQPSSTESMODLNKNREGKSRDLKDIKPSSAKDSFIENSNCYSGSSQPASTGTGOODSOHHGHSPAPSPLYOHNOQAALLNLASPOOGAIHYACLAGPAPPS TPASNTOSPNSFPAADOVTFTIHPSHOAPAUTPBPHAHNAHVQAHSVGSHPDTAHAPHMLMTTOPRCGCOALAALQSADLRPIPVTSTAHPMYTHPSVQAHHQQQL"
BASE COUNT	1144 a 1380 c 1014 g 943 t	
ORIGIN		
Query Match		100.0%; Score 20; DB 9; Length 4481;
Best Local Similarity		100.0% Pred. No. 1.2e+02;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Oy	1 gtgacgcaggaacaggagac 20	
Dd	833 GTGCCGAGAGCACGAGAC 814	
RESULT 8		
LOCUS	AC004085	231758 bp DNA linear HTG 06-NOV-2000
DEFINITION	Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered pieces.	
ACCSSION	AC004085	
VERSION	AC004085.6 GI:11079383	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	


```
REFERENCE 1 (bases 1 to 1209)
AUTHORS Schonlian,G.
JOURNAL Thesis (2001) Department of Institute of Microbiology and Hygiene,
Humbolt University, Berlin, Germany
REFERENCE 2 (bases 1 to 1209)
AUTHORS Merzlyak,E.M.
JOURNAL Direct Submission
TITLE Submitted (10-JAN-2001) Merzlyak E.M., Biology, Moscow State
UNIVERSITY University, Moscow, Vorob'evi Gori, 119899, RUSSIA
FEATURES
    source
        1. 1209
            /organism="Leishmania major"
            /strain="MHOM/ML87/NEI2"
            /db_xref="taxon:5664"
            1. 1209
            /gene="rnap"
            <1..>1209
            /gene="rnap"
            /function="polymerase"
            /codon_start=3
            /product="RNA polymerase II subunit"
            /protein_id="CAC2237.1"
            /translation="VKALEDVHASYGTVRNNOELLTOLAYGEDLDGARIENQAPFIP
            HMTNSEMDKRYREYNDEGSFSENNMGHYMDPVRDLSLADPQLIKLQEEFOLMRD
            RAMSKYLIDMEDKNKLKMLPVNVARLIONACTGTCKRSQVSNLPIYINVELOED
            LVQLPFSYHKDNGRFVNVLSQQRVERALTLFGIHLRQILGSKRVLKEYKLDKAFEY
            LLKEIRTKYQOSLITPGEIIGAIAOOSGEPATOMTLNTHNAGISSKVTLAVPELL
            ELLNYSRNORNASVAVCLIREYOKRNKAQAOOFLEYCTLIANTTTVOIITYPDPNPT
            VVAEDEMI RMEOAVNMADEDEPDPSPSPFIARLILNDLFI NDKRLNMKDKVKSIR
            QYDDTYMVQA"
BASE COUNT 285 a 353 c 369 g 199 t 3 others
ORIGIN
Query Match 95.0%; Score 19; DB 3; Length 1209;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtagccgagagacgagaga 19
|||||
Db 1008 GTGCCGACGAGCAGCAGAGA 1026

RESULT 10
AF009154 1268 bp DNA linear INV 02-NOV-1997
LOCUS Leishmania amazonensis RNA polymerase II large subunit gene,
DEFINITION partial cds.
ACCESSION AF009154
VERSION AF009154.1 GI:2581911
KEYWORDS
SOURCE Leishmania mexicana amazonensis.
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomat. lae;
Leishmania.
REFERENCE 1 (bases 1 to 1268)
AUTHORS Croan,D.G., Morrison,D.A. and Ellis,J.T.
JOURNAL Evolution of the genus Leishmania revealed by comparison of DNA and
TITLE RNA polymerase gene sequences
MOL. Biochem. Parasitol. 89 (2), 149-159 (1997)
JOURNAL 98030244
MEDLINE 2 (bases 1 to 1268)
REFERENCE Croan,D.G., Morrison,D.A. and Ellis,J.T.
AUTHORS Direct Submission
TITLE Submitted (18-JUN-1997) Cell & Molecular Biology, University of
JOURNAL Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia
FEATURES
    source
        1. 1268
            /organism="Leishmania mexicana amazonensis"
            /strain="MHOM/BR/73/LV78"
            /db_xref="taxon:5659"
            <1..>1268
mRNA
```

```
CDS
    /product="RNA polymerase II large subunit"
    <1..>1268
    /codon_start=2
    /product="RNA polymerase II large subunit"
    /protein_id="AAC47783.1"
    /db_xref="GI:2581912"
    /translation="TSDTCYLQKRLVKALEDVHASYDGTVRNANOELLTOLAYGEDLD
    GARISNOAFPIPHMTNSEMDKRYREYNDEGSFSENNMGHYMDPVRDLSLADPQSV
    LKLOEEFDLMDRAMSRVLIDMEDKNKLKMLPVNVARLIONARTTGTCKRSQVSNL
    PIVINRNVRELQEDLVOLFPFSYHKDNGRFVNVLSQQRVERALTLFGIHLRQILGSKR
    VLKEYKLDKAFEYLLKEIRTKYQOSLITPGEIIGAIAOOSGEPATOMTLNTHNAG
    ISSKVTGLGVPRLLELLNYSRNORNASVAVCLIREYOKRNKAQAOOFLEYCTLIANT
    TTVOIITYPDPNPTVVAEDEMI RMEOAVNMADEDEPDPSPSPFIARLILNDLFI
    NDKRLNMKDKVKSIRQYDDTYMVQANENDG"
BASE COUNT 302 a 364 c 392 g 210 t
ORIGIN
Query Match 95.0%; Score 19; DB 3; Length 1268;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtagccgagagacgagaga 19
|||||
Db 1049 GTGCCGACGAGCAGCAGAGA 1067

RESULT 11
AF009157 1268 bp DNA linear INV 02-NOV-1997
LOCUS Leishmania donovani RNA polymerase II large subunit gene, partial
DEFINITION cds.
ACCESSION AF009157
VERSION AF009157.1 GI:2581917
KEYWORDS
SOURCE Leishmania donovani.
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 1268)
AUTHORS Croan,D.G., Morrison,D.A. and Ellis,J.T.
JOURNAL Evolution of the genus Leishmania revealed by comparison of DNA and
TITLE RNA polymerase gene sequences
MOL. Biochem. Parasitol. 89 (2), 149-159 (1997)
JOURNAL 98030244
MEDLINE 2 (bases 1 to 1268)
REFERENCE Croan,D.G., Morrison,D.A. and Ellis,J.T.
AUTHORS Direct Submission
TITLE Submitted (18-JUN-1997) Cell & Molecular Biology, University of
JOURNAL Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia
FEATURES
    source
        1. 1268
            /organism="Leishmania donovani"
            /strain="MHOM/IN/80/DD8"
            /db_xref="taxon:5661"
            <1..>1268
            /product="RNA polymerase II large subunit"
            <1..>1268
            /product="RNA polymerase II large subunit"
            /codon_start=2
            /product="RNA polymerase II large subunit"
            /protein_id="AAC47786.1"
            /db_xref="GI:2581918"
            /translation="TSDTCYLQKRLVKALEDVHASYDGTVRNANOELLTOLAYGEDLD
            GARISNOAFPIPHMTNSEMDKRYREYNDEGSFSENNMGHYMDPVRDLSLADPQSV
            LKLOEEFDLMDRAMSRVLIDMEDKNKLKMLPVNVARLIONARTTGTCKRSQVSNL
            PIVINRNVRELQEDLVOLFPFSYHKDNGRFVNVLSQQRVERALTLFGIHLRQILGSKR
            VLKEYKLDKAFEYLLKEIRTKYQOSLITPGEIIGAIAOOSGEPATOMTLNTHNAG
            ISSKVTGLGVPRLLELLNYSRNORNASVAVCLIREYOKRNKAQAOOFLEYCTLIANT
            TTVOIITYPDPNPTVVAEDEMI RMEOAVNMADEDEPDPSPSPFIARLILNDLFI
            NDKRLNMKDKVKSIRQYDDTYMVQANENDG"
BASE COUNT 301 a 372 c 392 g 203 t
ORIGIN
```

Query Match 95.0%; Score 19; DB 3; Length 1268;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgacgagagcagagaga 19
 |||||||
 Db 1049 GTGGCCGAGCAGCAGAGA 1067

RESULT 12

AF009164

LOCUS 1268 bp DNA linear INV 02-NOV-1997
 DEFINITION Leishmania mexicana RNA polymerase II large subunit gene, partial
 cds.

ACCESSION AF009164
 VERSION AF009164.1 GI:2581931

KEYWORDS
 ORGANISM Leishmania mexicana.
 Leishmania mexicana
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE 1 (bases 1 to 1268)
 Croan,D.G., Morrison,D.A. and Ellis,J.T.
 Evolution of the genus Leishmania revealed by comparison of DNA and
 RNA polymerase gene sequences
 Mol. Biochem. Parasitol. 89 (2), 149-159 (1997)

JOURNAL 2 (bases 1 to 1268)
 Croan,D.G., Morrison,D.A. and Ellis,J.T.
 Direct Submission
 Submitted (18-JUN-1997) Cell & Molecular Biology, University of
 Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia

FEATURES
 source
 1..1268
 /organism="Leishmania mexicana"
 /strain="MNYC/B2/G2/M379"
 /db_xref="taxon:5665"

FEATURES

mRNA

CDS

/product="RNA polymerase II large subunit"
 <1..>1268
 /codon_start=2
 /product="RNA polymerase II large subunit"
 /protein_id="AAC47793.1"
 /db_xref="GI:2581932"

/translation="TSDTGYLQRLKLVKALVDYDGTVRNANOELLQALVGEGLD
 GARIEGNOAFPIPHMTNSEMAEKRYREYEDSGFSFENGGHYMDPFVDSLLRDPQSV
 LKLOEFQDLMDRAMSLVLDIMEDKNLKNMLPVNVARLIONARTTGRKRSQVSNLN
 PTTVINRPRELEDVQLFPRSHKDYNGRFVYVLSQOVERALTFTGLHRLQILGSKR
 VLKRYLNDKAFEYLKELRTKYQSLITPGELTIGALIAQSCGEPATOWTLNTPHNAG
 ISSKNVTLGVPRLLELNLVSKNQRNASVAVCLIRGYQRNKAQEOQFIETCLANIT
 TTVQIYDPPDPRNTYVAEDEMI RMEQAVMNEDEDEPDPAEQPPSPFIARLLIDNDFN
 DKRLNKDVKSARVQDYDYYVAQAMENNG"

BASE COUNT 302 a 364 c 395 g 207 t
 ORIGIN

Query Match 95.0%; Score 19; DB 3; Length 1268;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgacgagagcagagaga 19
 |||||||
 Db 1049 GTGGCCGAGCAGCAGAGA 1067

RESULT 13

AF009167

LOCUS 1268 bp DNA linear INV 02-NOV-1997
 DEFINITION Leishmania tropica RNA polymerase II large subunit gene, partial
 cds.

ACCESSION AF009167
 VERSION AF009167.1 GI:2581937

KEYWORDS
 SOURCE
 ORGANISM

Leishmania tropica.
 Leishmania tropica
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE 1 (bases 1 to 1268)
 Croan,D.G., Morrison,D.A. and Ellis,J.T.
 Evolution of the genus Leishmania revealed by comparison of DNA and
 RNA polymerase gene sequences
 Mol. Biochem. Parasitol. 89 (2), 149-159 (1997)

JOURNAL

2 (bases 1 to 1268)
 Croan,D.G., Morrison,D.A. and Ellis,J.T.
 Direct Submission
 Submitted (18-JUN-1997) Cell & Molecular Biology, University of
 Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia

FEATURES
 source
 1..1268
 /organism="Leishmania tropica"
 /strain="MNYC/5U/58/Strain-OD"
 /db_xref="taxon:5666"

mRNA

CDS

/product="RNA polymerase II large subunit"
 <1..>1268
 /codon_start=2
 /product="RNA polymerase II large subunit"
 /protein_id="AAC47796.1"
 /db_xref="GI:2581938"

/translation="TSDTGYLQRLKLVKALVDYDGTVRNANOELLQALVGEGLD
 GARIEGNOAFPIPHMTNSEMAEKRYREYEDSGFSFENGGHYMDPFVDSLLRDPQSV
 LKLOEFQDLMDRAMSLVLDIMEDKNLKNMLPVNVARLIONARTTGRKRSQVSNLN
 PTTVINRPRELEDVQLFPRSHKDYNGRFVYVLSQOVERALTFTGLHRLQILGSKR
 VLKRYLNDKAFEYLKELRTKYQSLITPGELTIGALIAQSCGEPATOWTLNTPHNAG
 ISSKNVTLGVPRLLELNLVSKNQRNASVAVCLIRGYQRNKAQEOQFIETCLANIT
 TTVQIYDPPDPRNTYVAEDEMI RMEQAVMNEDEDEPDPAEQPPSPFIARLLIDNDFN
 DKRLNKDVKSARVQDYDYYVAQAMENNG"

BASE COUNT 296 a 373 c 398 g 201 t
 ORIGIN

Query Match 95.0%; Score 19; DB 3; Length 1268;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgacgagagcagagaga 19
 |||||||
 Db 1049 GTGGCCGAGCAGCAGAGA 1067

RESULT 14

AF126254

LOCUS 5500 bp DNA linear INV 13-MAR-2001
 DEFINITION Leishmania donovani RNA polymerase II largest subunit (RNP11L3)
 gene, complete cds.

ACCESSION AF126254
 VERSION AF126254.2 GI:13310931

KEYWORDS

Leishmania donovani.
 Leishmania donovani.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE 1 (bases 1 to 5500)
 Dasgupta,A., Sharma,S., Das,A. and Majumder,H.K.
 Molecular cloning and characterization of RNA polymerase II largest
 subunit gene of Leishmania donovani

JOURNAL

2 (bases 1 to 5500)
 Dasgupta,A., Sharma,S., Das,A. and Majumder,H.K.
 Direct Submission
 Submitted (05-FEB-1999) Molecular Parasitology, Indian Institute of
 Chemical Biology, 4, Raja S.C. Mullick Road, Calcutta, West Bengal
 700 032, India

REFERENCE 3 (bases 1 to 5500)

AUTHORS Dasgupta, A., Sharma, S., Das, A. and Majumder, H.K.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2001) Molecular Parasitology, Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Calcutta, West Bengal 700 032, India
 REMARK Sequence update by submitter
 COMMENT On Mar 13, 2001 this sequence version replaced gi:7108488.
 FEATURES
 source location/Qualifiers
 1..5500
 /organism="Leishmania donovani"
 /strain="D1700"
 /db_xref="taxon:5661"
 100..5097
 /gene="RNP11LS"
 100..5097
 /gene="RNP11LS"
 /codon_start=1
 /product="RNA polymerase II largest subunit"
 /protein_id="AAF36438.2"
 /db_xref="GI:13310932"

gene
 CDS

BASE COUNT 1145 a 1699 c 1680 g 976 t
 ORIGIN

Query Match 95.0%; Score 19; DB 3; Length 5500;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtggcgaagagagagaga 19
 |||||
 Db 3577 GTGGCGAGAGAGAGA 3595

RESULT 15
 AF009163
 LOCUS 7236 bp DNA linear INV 25-JUL-2001
 DEFINITION Leishmania major RNA polymerase II large subunit (RNP11LS) gene,
 complete cds; and calreticulin (CLR) gene, partial cds.
 ACCSSION AF009163 AF120931
 VERSION AF009163.2 GI:5263287
 KEYWORDS
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 7236)
 AUTHORS Croan,D.G., Morrison,D.A. and Ellis,J.T.
 TITLE Evolution of the genus Leishmania revealed by comparison of DNA and RNA polymerase gene sequences
 JOURNAL Mol. Biochem. Parasitol. 89 (2), 149-159 (1997)
 MEDLINE 98030244
 REFERENCE 2 (bases 1 to 7236)
 AUTHORS Croan,D.G. and Ellis,J.
 TITLE The Leishmania major RNA polymerase II largest subunit lacks a carboxy-terminus heptad repeat structure and its encoding gene is linked with the calreticulin gene
 JOURNAL Proteist 151 (1), 57-68 (2000)
 MEDLINE 20353029
 PUBMED 10896133
 REFERENCE 3 (bases 2754 to 4021)
 AUTHORS Croan,D.G., Morrison,D.A. and Ellis,J.T.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-1997) Cell & Molecular Biology, University of Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia
 4 (bases 1 to 7236)
 AUTHORS Croan,D.G. and Ellis,J.T.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-1999) Cell & Molecular Biology, University of Technology Sydney, Westbourne St., Gore Hill, NSW 2065, Australia
 REMARK Sequence update by submitter
 COMMENT On Jun 29, 1999 this sequence version replaced gi:2581929.
 FEATURES
 source location/Qualifiers
 1..7236
 /organism="Leishmania major"
 /strain="MHOM/SU/73/5-ASKH"
 /db_xref="taxon:5664"
 /dev_stage="promastigote"
 <325..>5316
 /gene="RNP11LS"
 /product="RNA polymerase II large subunit"
 <325..>5316
 /gene="RNP11LS"
 325..5316
 /gene="RNP11LS"
 /codon_start=1
 /product="RNA polymerase II large subunit"
 /protein_id="AAC47792.2"
 /db_xref="GI:5263288"

gene
 CDS

translation="MSGGAALPSPOMPLQKVHEVOFEVEFEKRAQIKAYAKCIEHAASY
 ASTASGCAASTTCAMAROTSTIACGCRHPCGCHGYELAEIPINIVPEVYL
 IALCKVCYKGLAMTNDPSEKKTAHLQGLRMLVMVAKLIGSCVKRSKDLQGEKG
 GPPAARHPVGYIPGLOIVTQEDDYVMAENARQVLDVSDSALIMGFRRCHP
 RDLDLVLP1PPOVRSIYLGLMGVDLTQIMSLVRNQLRKDKSGYKAAVDR
 SRALQEHVATPEVNNASTYYKPAKVGDTKLKSLTERLKGKGRGLNLMGKRVFSA
 RYITGDNDVDEGVPEVSAMTLTEPEEVNINIKRLTEFQRTTYSANYITRPN
 GNYTRALVKDRAVHLVDGVYERHVDVLEFQDYLTFNRQPLHMSMGHRAVLYNT
 RLNTSCPTPYNADFDGDEKMLHVPQSLTKAEILMMVPRKFTSPKSAQCKIVQ
 DSVLGSYKLEKDTLDKYFIARPLMLDLMOPLPILKPRLMYTGKQIFSLIPEV
 NHPASHPDKPPPHNDKIMIOGQLLGAITLGGVGAAGSLIHVIFENKSDYKAK
 FINGQVITTYENYCAFSVGVODTADATTLKEMNNVJHKTKOSVEKIGAAANNQIT
 TRAGKSLQSFADVNSALNCKREKAAKALSNVRRTSFKYMLIAGSGSDNLNCO
 IAYFGQONVAGSRIPGFRRTLPHEMLDDYSETSRGATRGVYGLDLPHEHYFTM
 AGREGILTRAVKTSPTGYLQKRYALKEVDYHAYSDCTVRNANDELQIAYGDDGDA
 RLEGNOAFPIPHMTNSEMDKRYETENDSGSENNMGHMDPFVDSILRPOSVLK
 LOEFBOLMKDRAVSLVIDMEDKNLKNMLPVNVARLIONATTNGKSOVSNLPI
 TVINRRELQEDLVOLPSTYHKNQIRFVNVLSQOEVERALTLFGHLQILGSKRVL
 KEYKLNDAFEYLKEIRTKYQOGLTPEIGALIAQSGCEPATOMTLNTEFNAGIS
 SKNVTGLVPRLELLLVNKNORNASVACLIREYOKRNKAQEQOIEVCTLANITTT
 VOIITPPDPRNTYVADEDEKIRNEQAVNNAEDEDEPAEQPSFIARLILDNDFDK
 RLNMKDVKSALROVDITYVOAMENDGORTVLRPKCTGADSPALTRKAAVQLQAS
 VHLRIGPIKTKTLKGNTRFVDPENAGIKNESMWVDEGTALQRIPIGVANSEGN
 IIDFSKTSNKKIPEVTVVIGIEAARKKLSELBEAYLAGININRYHITLDTMGOR
 GYLMAVSRGINSSETSGPLMRCSFEETVVLMTAAFGKEDPRGVSASLYGNQAR
 IGTLGDLILDMNSKIQHVPLDKATEARTSNVHTDASVAPGSLQSGHGLPSTV
 HENSSLAAGASVYPRHERAGMTGGIPIEASVQFSSALPTVMTALAEHDRVQO
 PPPLMCFHVLGVGAERESLDLELSYHLSQVAPSTAYTATYAPMAASGMRPQATOLS
 VBGSGSLPYPEASNGERFSGISGGAASRASAVYDPTQOPQSEFEFSPTREDEEP"

mRNA <6364. .>7236
/gene="CLT"
/product="calreticulin"
gene <6364. .>7236
/gene="CLT"
CDS 6364. .>7236
/gene="CLT"
/note="calcium binding protein; molecular chaperone;
RNA-binding phosphoprotein"
/codon_start=1
/product="calreticulin"
/protein_id="AAD4141.1"
/db_xref="GI:5263289"
/translation="MLAIVGVLCVYVVOAEIFFEHEEFTMDGWOSEHTSDYGV
ALSYGAIHVDAERKEDQIKMEDAKFYAVSKILPKAVSNDKPLIVSESVKNBOKLTG
GTYLRFSELDOKDLHGESATYLMFGPDICGSKNRLOFILSTNGTNHLMKLSTPKTD
TATHVYTVETIAPNNYVQLYVDGMHIQEGSLEEDMDLPPKTI SDPTDKKPADWVDDMM
MDPSPDTKPEADWDEPATITDSEAVKPVWDSDSDGVWEAPKVPNPYPGAMKPRRIH
NPDYKGEAARQIEL"

BASE COUNT 1507 a 2174 c 2188 g 1357 t 10 others
ORIGIN

Query Match 95.0%; Score 19; DB 3; Length 7236;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtgcccagagacgagaga 19
|||||
Db 3802 GTGCCCGAGGACGAGAGA 3820

Search completed: August 14, 2002, 21:47:56
Job time: 13494 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:50:47 : Search time 203.42 Seconds
(without alignments)
24.150 Million cell updates/sec

Title: US-09-707-919-2
Perfect score: 20
Sequence: 1 gtggccgagcagcagcagac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	100.0	205	4	US-09-043-303-3 Sequence 3, Appl1
C 2	20	100.0	623	4	US-09-043-303-5 Sequence 5, Appl1
C 3	20	100.0	4481	4	US-09-041-886-18 Sequence 18, Appl1
C 4	16.8	84.0	2634	3	US-08-911-853-30 Sequence 30, Appl1
C 5	16.8	84.0	2634	4	US-09-479-409-30 Sequence 30, Appl1
C 6	16.8	84.0	2634	4	US-09-479-453-30 Sequence 29, Appl1
C 7	16.8	84.0	17612	3	US-08-911-853-29 Sequence 29, Appl1
C 8	16.8	84.0	17612	4	US-09-479-409-29 Sequence 29, Appl1
C 9	16.8	84.0	17612	4	US-09-479-453-29 Sequence 29, Appl1
C 10	16.4	82.0	1543	3	US-09-339-775-1 Sequence 1, Appl1
C 11	16	80.0	28958	1	US-08-258-261B-6 Sequence 6, Appl1
C 12	16	80.0	28958	1	US-08-456-837-6 Sequence 6, Appl1
C 13	16	80.0	28958	1	US-08-457-342-6 Sequence 6, Appl1
C 14	16	80.0	28958	1	US-08-458-076A-6 Sequence 6, Appl1
C 15	16	80.0	28958	1	US-08-458-076A-6 Sequence 6, Appl1
C 16	16	80.0	28958	1	US-08-764-233A-4 Sequence 6, Appl1
C 17	16	80.0	28958	1	US-08-457-335A-6 Sequence 6, Appl1
C 18	16	80.0	28958	1	US-08-729-214-6 Sequence 6, Appl1
C 19	16	80.0	28958	3	US-09-028-934-6 Sequence 6, Appl1
C 20	16	80.0	49377	1	US-08-764-233A-1 Sequence 1, Appl1
C 21	15.8	79.0	65	2	US-08-707-237A-27 Sequence 27, Appl1
C 22	15.8	79.0	1910	1	US-08-247-902A-1 Sequence 1, Appl1
C 23	15.8	79.0	1910	5	PCT-US93-10541-1 Sequence 1, Appl1
C 24	15.8	79.0	3097	4	US-09-282-147-38 Sequence 38, Appl1
C 25	15.8	79.0	3257	5	PCT-US91-09784-1 Sequence 1, Appl1
C 26	15.8	79.0	4403765	4	US-09-103-840A-2 Sequence 2, Appl1
C 27	15.8	79.0	4411529	4	US-09-103-840A-1 Sequence 1, Appl1

28	15.4	77.0	2104	4	US-09-313-930-1	Sequence 1, Appl1
29	15.4	77.0	2721	6	5215881-2	Patent No. 5215881
30	15.4	77.0	6601	3	US-09-356-952-10	Sequence 10, Appl1
31	15.4	77.0	8438	1	US-07-945-283-1	Sequence 1, Appl1
32	15.2	76.0	1720	4	US-09-227-357-139	Sequence 139, App
33	15.2	76.0	1748	3	US-09-100-730-1	Sequence 1, Appl1
34	15.2	76.0	1869	3	US-08-952-967-7	Sequence 7, Appl1
35	15.2	76.0	1947	1	US-07-998-972A-2	Sequence 2, Appl1
36	15.2	76.0	1947	1	US-08-463-953-2	Sequence 2, Appl1
37	15.2	76.0	1947	1	US-08-462-261-2	Sequence 2, Appl1
38	15.2	76.0	1947	2	US-08-479-733A-24	Sequence 24, Appl1
39	15.2	76.0	1947	2	US-08-487-427-24	Sequence 24, Appl1
40	15.2	76.0	1947	3	US-08-479-727A-24	Sequence 24, Appl1
41	15.2	76.0	1947	3	US-08-482-369A-24	Sequence 24, Appl1
42	15.2	76.0	1947	5	PCT-US92-11357-2	Sequence 2, Appl1
43	15.2	76.0	1947	5	PCT-US95-07439-24	Sequence 24, Appl1
44	15.2	76.0	1988	1	US-07-750-080A-15	Sequence 15, Appl1
45	15.2	76.0	1988	3	US-08-651-472-15	Sequence 15, Appl1

ALIGNMENTS

```
RESULT 1
US-09-043-303-3/c
; Sequence 3, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 205
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(174)
US-09-043-303-3

Query Match      100.0%; Score 20; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  gtggccgagcagcagcagac 20
Db      110 GTGCCGAGCAGCAGCAGAC 91

RESULT 2
US-09-043-303-5/c
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
```

```
SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(583)
; FEATURE:
; OTHER INFORMATION: Tsp-2
US-09-043-303-5

Query Match          100.0%; Score 20; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 gtggccgagagagagagac 20
        |||
Db      528 GTGCCGAGAGAGAGAGAC 509

RESULT 3
US-09-041-886-18/c
; Sequence 18, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; TITLE OF INVENTION: Prooptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..4099
US-09-041-886-18

Query Match          100.0%; Score 20; DB 4; Length 4481;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 gtggccgagagagagagac 20
```

```
|||||
Db      833 GTGCCGAGAGAGAGAGAC 814

RESULT 4
US-08-911-853-30/c
; Sequence 30, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gjsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-911-853-30

Query Match          84.0%; Score 16.8; DB 3; Length 2634;
Best Local Similarity 90.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 gtggccgagagagagagac 20
        |||
Db      1784 GTGCCGAGAGAGAGAGAC 1765

RESULT 5
US-09-479-409-30/c
; Sequence 30, Application US/09479409
; Patent No. 6225106
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gjsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
```



```

;
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,409
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-479-409-30

```

```

Query Match      84.0%; Score 16.8; DB 4; Length 2634;
Best Local Similarity 90.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```
OY      1 gtggccgagagcagagac 20
        |||||||||||||
Db      1784 GTGCCGAGAGCAGCGTGAC 1765

```

```

RESULT      6
US-09-479-453-30/c
; Sequence 30, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 30:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-479-453-30

```

```

Query Match      84.0%; Score 16.8; DB 4; Length 2634;
Best Local Similarity 90.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```
OY      1 gtggccgagagcagagac 20
        |||||||||||||
Db      1784 GTGCCGAGAGCAGCGTGAC 1765

```

```

RESULT      7
US-08-911-853-29/c
; Sequence 29, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-911-853-29

```

```

Query Match      84.0%; Score 16.8; DB 3; Length 17612;
Best Local Similarity 90.0%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```
OY      1 gtggccgagagcagagac 20
        |||||||||||||
Db      6199 GTGCCGAGAGCAGCGTGAC 6180

```

```

RESULT      8
US-09-479-409-29/c
; Sequence 29, Application US/09479409

```

```
Patent No. 6225106
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,409
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 17612 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-409-29

Query Match      84.0%; Score 16.8; DB 4; Length 17612;
Best Local Similarity 90.0%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtggccgagacgagagac 20
DB 6199 GTGCCGAGACGACGTGAC 6180

RESULT 9
US-09-479-453-29/c
Sequence 29, Application US/09479453
Patent No. 6313283
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/09/479,453
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 17612 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-453-29
```

```
Query Match      84.0%; Score 16.8; DB 4; Length 17612;
Best Local Similarity 90.0%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtggccgagacgagagac 20
DB 6199 GTGCCGAGACGACGTGAC 6180
```

```
RESULT 10
US-09-339-775-1
Sequence 1, Application US/09339775
Patent No. 6063626
GENERAL INFORMATION:
APPLICANT: Lex M. Cowser
TITLE OF INVENTION: ANTISENSE MODULATION OF G-ALPHA-13 EXPRESSION
FILE REFERENCE: RTS-0069
CURRENT APPLICATION NUMBER: US/09/339,775
CURRENT FILING DATE: 1999-06-24
NUMBER OF SEQ ID NOS: 47
SEQ ID NO 1
LENGTH: 1543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (9)..(1073)
US-09-339-775-1
```

```
Query Match      82.0%; Score 16.4; DB 3; Length 1543;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 ttggccgagacgagagaga 19
DB 709 ttgctgagagcagagaga 726
```

```
RESULT 11
US-08-258-261B-6/c
Sequence 6, Application US/08258261B
Patent No. 5639949
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
```

```

? APPLICANT: Ukes, Scott Joseph
? TITLE OF INVENTION: Genes for the synthesis of
? TITLE OF INVENTION: antipathogenic substances
? NUMBER OF SEQUENCES: 22
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Ciba-Geigy Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/258,261B
? FILING DATE: 08-JUN-1994
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/457,205
? FILING DATE: 01-JUN-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Elmer, James Scott
? REGISTRATION NUMBER: 36,129
? REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8614
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 28958 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
?
US-08-258-261B-6
?
Query Match 80.0%; Score 16; DB 1; Length 28958;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtggcgagagcgagg 16
|||||
Db 23640 GTGGCCGAGCAGCAGG 23625

RESULT 12
US-08-456-837-6/C
? Sequence 6, Application US/08456837
? Patent No. 5643774
? GENERAL INFORMATION:
? APPLICANT: Schupp, Thomas
? APPLICANT: Ligon, James M.
? APPLICANT: Beck, James Joseph
? APPLICANT: Hill, Dwight Steven
? APPLICANT: Ryals, John Andrew
? APPLICANT: Gaffney, Thomas Deane
? APPLICANT: Lam, Stephen Ting
? APPLICANT: Hammer, Phillip E.
? APPLICANT: Ukes, Scott Joseph
? TITLE OF INVENTION: Genes for the synthesis of
? TITLE OF INVENTION: antipathogenic substances
? NUMBER OF SEQUENCES: 22
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Ciba-Geigy Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
```

```

? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/456,837
? FILING DATE: 01-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/457,205
? FILING DATE: 01-JUN-1995
? APPLICATION NUMBER: 08/258,261
? FILING DATE: 08-JUN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Elmer, James Scott
? REGISTRATION NUMBER: 36,129
? REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8614
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 28958 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
?
US-08-456-837-6
?
Query Match 80.0%; Score 16; DB 1; Length 28958;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtggcgagagcgagg 16
|||||
Db 23640 GTGGCCGAGCAGCAGG 23625

RESULT 13
US-08-457-342-6/C
? Sequence 6, Application US/08457342
? Patent No. 5662898
? GENERAL INFORMATION:
? APPLICANT: Schupp, Thomas
? APPLICANT: Ligon, James M.
? APPLICANT: Beck, James Joseph
? APPLICANT: Hill, Dwight Steven
? APPLICANT: Ryals, John Andrew
? APPLICANT: Gaffney, Thomas Deane
? APPLICANT: Lam, Stephen Ting
? APPLICANT: Hammer, Phillip E.
? APPLICANT: Ukes, Scott Joseph
? TITLE OF INVENTION: Genes for the synthesis of
? TITLE OF INVENTION: antipathogenic substances
? NUMBER OF SEQUENCES: 22
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Ciba-Geigy Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
;
US-08-457-342-6

Query Match      80.0%; Score 16; DB 1; Length 28958;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gtggccgagcagcag 16
Db      23640 GTGCCCGAGCAGCAGC 23625

RESULT 14
US-08-457-646A-6/c
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
```

```

; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
;
US-08-457-646A-6

Query Match      80.0%; Score 16; DB 1; Length 28958;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gtggccgagcagcag 16
Db      23640 GTGCCCGAGCAGCAGC 23625

RESULT 15
US-08-458-076A-6/c
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: 919-541-8614
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 28958 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-458-076A-6

Query Match 80.0%; Score 16; DB 1; Length 28958;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTGCGGAGGACGAGG 16
 |||||
 DB 23640 GTGCGGAGGACGAGG 23625

Search completed: August 14, 2002, 21:51:00
 Job time: 13493 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:00 ; Search time 7749.14 Seconds
(without alignments)

34.835 Million cell updates/sec

Title: US-09-707-919-2
Perfect score: 20
Sequence: 1 gtggcggagcgagcgagac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	357	9	AA065280 e01500t T
2	20	100.0	405	9	AA065279 e10500r T
3	20	100.0	839	9	AL538923 AL538923
4	20	100.0	1100	10	BM455214 AGENCOURT
5	19	95.0	638	12	BH019669 BH019669
6	19	95.0	654	12	BH019726 BH019726
7	18.4	92.0	401	10	U80749 U80749
8	18.4	92.0	482	9	AL039573 AL039573
9	18.4	92.0	500	10	B1547486 B1547486
10	18.4	92.0	567	9	AI513242 AI513242
11	18.4	92.0	632	10	BE585103 BE585103
12	18.4	92.0	660	9	AM845240 AM845240
13	18.4	92.0	854	9	AU124593 AU124593
14	18.4	92.0	874	10	BG387406 BG387406
15	18.4	92.0	910	10	BF204258 BF204258
16	18.4	92.0	932	10	BE908252 BE908252
17	18.4	92.0	936	10	BF984116 BF984116

c	18	18.4	92.0	951	10	BE780737	BE780737	601469609
c	19	18.4	92.0	2530	11	BC015228	BC015228	Hom sapi
c	20	18.4	92.0	2676	11	BC013955	BC013955	Hom sapi
c	21	18.4	92.0	277	11	BC635283	BC635283	AT31811.5
c	22	18.4	92.0	473	9	AA696695	AA696695	GM08174.5
c	23	18.4	92.0	496	9	AA695428	AA695428	GM02816.5
c	24	18.4	92.0	505	10	BC633872	BC633872	AT29828.5
c	25	18.4	92.0	561	9	AI532761	AI532761	SD04367.5
c	26	18.4	92.0	569	10	BI653203	BI653203	SD16475.5
c	27	18.4	92.0	572	10	BC641226	BC641226	SD12516.5
c	28	18.4	92.0	587	9	AA803325	AA803325	GM10575.3
c	29	18.4	92.0	591	10	BI352858	BI352858	GM21085.5
c	30	18.4	92.0	594	10	BE978189	BE978189	bs75c01.y
c	31	18.4	92.0	616	10	BI352748	BI352748	GM20586.5
c	32	18.4	92.0	620	10	BI230134	BI230134	GM14980.5
c	33	17.4	87.0	150	9	AM935384	AM935384	CM3-DT000
c	34	17.4	87.0	262	10	BE386688	BE386688	946012F12
c	35	17.4	87.0	448	9	AU077604	AU077604	AU077604
c	36	17.4	87.0	455	10	BI482446	BI482446	RE65141.5
c	37	17.4	87.0	473	12	CNS0451L	AL275538	Tetradon
c	38	17.4	87.0	475	12	BH411845	BH411845	1007024C0
c	39	17.4	87.0	499	12	BH411846	BH411846	1007024C0
c	40	17.4	87.0	500	12	BH411848	BH411848	1007024C0
c	41	17.4	87.0	536	9	AI134913	AI134913	GH12493.5
c	42	17.4	87.0	543	10	BE605309	BE605309	WHE2331_B
c	43	17.4	87.0	551	10	BI354102	BI354102	GM26304.5
c	44	17.4	87.0	551	10	BE470763	BE470763	WHE0281.H
c	45	17.4	87.0	554	10	BF277605	BF277605	GA_Eb003

ALIGNMENTS

RESULT 1
LOCUS AA065280 357 bp mRNA linear EST 31-DEC-1996
DEFINITION e01500t Testis 5 Homo sapiens cDNA clone e01500 5' end, mRNA
sequence.
ACCESSION AA065280
VERSION AA065280.1 GI:1929280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Guellaen, G.
TITLE Unpublished (1996)
JOURNAL Unpublished (1996)
COMMENT Contact: Guellaen G
Unite INSERM 99
INSERM
Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)14980908
Email: guellaen@infobiogen.fr
This sequence derives from a clone which was selected from the cDNA
library - Testis 5 - using a repeat of 14 CAG as probe
Seq primer: T7.
Location/Qualifiers
1. 357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="e01500"
/clone_lib="Testis 5"
/note="Vector: pSPORT1; Site_1: MluI; Site_2: NotI; mRNA
was prepared from human testis of a 27 years old man. cDNA
was prepared using a 15mer oligo dT anchored by two
degenerated bases at its 3' end and containing a NotI site
at its 5' end. The cDNA was cloned between SalI and NotI
sites of pSPORT1. The MluI-SalI fragment come from the
adaptor used for the cloning. The 5' end is at the NotI
site. cDNA corresponding to abundant species were

```

BASE COUNT      44 a      eliminated from this library.*
ORIGIN           111 c      127 g      65 t      10 others

Query Match      100.0%; Score 20; DB 9; Length 357;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  gtggccgagagcagagagac 20
        |||
        69  GTGCCGAGCAGCAGCAGAC 88

RESULT 2
AA065279      405 bp      mRNA      linear      EST 31-DEC-1996
LOCUS      e10500r Testis 5 Homo sapiens cDNA clone e10500 3' end, mRNA
DEFINITION      sequence.
ACCESSION      AA065279
VERSION      AA065279.1 GI:1929279
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
TITLE      1 (bases 1 to 405)
JOURNAL      Guellaen,G. Unpublished (1996)
COMMENT      Unpublished (1996)
              Contact: Guellaen G
              Unite INSERM 99
              INSERM
              Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
              Tel: (33)149813530
              Fax: (33)148980908
              Email: guellaen@infobiogen.fr
              This sequence derives from a clone which was selected from the cDNA
              library - Testis 5 - using a repeat of 14 CAG as probe
              Seq primer: M13 reverse.
              Location/Qualifiers
                1. 405
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="Testis 5"
                /note="Vector: pSPORT1; site_1: MluI; site_2: NotI; mRNA
                was prepared from human testis of a 27 years old man. cDNA
                degenerated bases at its 3' end and containing a NotI site
                at its 5' end. The cDNA was cloned between SalI and NotI
                sites of pSPORT1. The MluI-SalI fragment come from the
                adaptor used for the cloning. The 3' end is at the NotI
                site. cDNA corresponding to abundant species were
                eliminated from this library."

BASE COUNT      66 a      122 c      136 g      77 t      4 others
ORIGIN

Query Match      100.0%; Score 20; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  gtggccgagagcagagagac 20
        |||
        69  GTGCCGAGCAGCAGCAGAC 88

RESULT 3
AL538923      839 bp      mRNA      linear      EST 16-FEB-2001
LOCUS      AL538923 LTI_FL013_Fbrn1 Homo sapiens cDNA clone CS0DF030YB07 5
DEFINITION      prime, mRNA sequence.

```

```

ACCESSION      AL538923
VERSION      AL538923.1 GI:12867670
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
TITLE      1 (bases 1 to 839)
JOURNAL      Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
              Full-length cDNA libraries and normalization
              Unpublished (2001)
COMMENT      Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 Evry cedex - France
              Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
  source
    1. 839
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="CS0DF030YB07"
    /clone_lib="LTI_FL013_Fbrn1"
    /dev_stage="pooled tissue from post conception fetuses (20
    week, 24 week and 26 week)"
    /lab_host="DH10B"
    /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
    cDNA was primed with a NotI-oligo(dt) primer. Five prime
    end enriched, double-stranded cDNA was digested with Not I
    and cloned into the Not I and Eco RV sites of the
    pCMVSPORT 6 vector. Library was constructed by life
    Technologies. Contact : Feng Liang Life Technologies, a
    division of Invitrogen 9800 Medical Center Drive Rockville
    , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
    fliang@lifetechn.com URL :
    http://fulllength.invitrogen.com"

BASE COUNT      77 a      295 c      294 g      154 t      19 others
ORIGIN

Query Match      100.0%; Score 20; DB 9; Length 839;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  gtggccgagagcagagagac 20
        |||
        90  GTGCCGAGCAGCAGCAGAC 71

RESULT 4
BM455214      1100 bp      mRNA      linear      EST 05-FEB-2002
LOCUS      AGENCOURT_6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
DEFINITION      5', mRNA sequence.
ACCESSION      BM455214
VERSION      BM455214.1 GI:18504254
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
TITLE      NIH-MGC http://mgc.ncl.nih.gov/
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-r@mail.nih.gov
              Tissue Procurement: Lou Staudt
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM12134 row: k column: 12

```

High quality sequence stop: 623.

FEATURES

Location/Qualifiers

1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5500163"
/clone_id="NIH_MGC_85"
/issue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 240 a 329 c 306 g 219 t 6 others

ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 8.7e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagagcagagagac 20
|||||
Db 457 GTGCCGAGCAGCAGAGAC 438

RESULT 5 638 bp DNA linear GSS 25-MAY-2001
BH019669/c

LOCUS L456k.d.Hygt3.1 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION BH019669
Leishmania major genomic clone L456k, DNA sequence.

ACCESSION BH019669
VERSION BH019669.1 GI:14199118

KEYWORDS GSS.
Leishmania major.

SOURCE Leishmania major.
ORGANISM Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 638)

AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Kiackling,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal,
,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.

TITLE Leishmania major Friedlin Cosmid End Sequences
JOURNAL Unpublished (2000)

COMMENT Other GSS: L456k.d.Hygt7a.1

Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@bri.org

Seq primer: Hygt3
Class: cosmid ends.

FEATURES Location/Qualifiers

1..638
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L456k"
/clone_id="Leishmania major Friedlin Cosmid Genomic Library"

/lab_host="E. coli ED8767"
/note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sau3AI, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivens et al., Genomics Research, 8:135-145 (1998). The cLHYG vector (ACC. No. CVU59231) is described in Ryan et al., Gene, 131:145-150 (1993)"

BASE COUNT 102 a 192 c 193 g 151 t

ORIGIN

Query Match 95.0%; Score 19; DB 12; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagagcagagaga 19
|||||
Db 190 GTGCCGAGCAGCAGAGAGA 172

RESULT 6 654 bp DNA linear GSS 25-MAY-2001
BH019726/c

LOCUS L473k.d.Hygt3.1 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION BH019726
Leishmania major genomic clone L473k, DNA sequence.

ACCESSION BH019726
VERSION BH019726.1 GI:14199238

KEYWORDS GSS.
Leishmania major.

SOURCE Leishmania major.
ORGANISM Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 654)

AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Kiackling,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal,
,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.

TITLE Leishmania major Friedlin Cosmid End Sequences
JOURNAL Unpublished (2000)

COMMENT Other GSS: L473k.d.Hygt7a.1

Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@bri.org

Seq primer: Hygt3
Class: cosmid ends.

FEATURES Location/Qualifiers

1..654
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L473k"
/clone_id="Leishmania major Friedlin Cosmid Genomic Library"

/lab_host="E. coli ED8767"
/note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sau3AI, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Ivens et al., Genomics Research, 8:135-145 (1998). The cLHYG vector (ACC. No. CVU59231) is described in Ryan et al., Gene, 131:145-150 (1993)"

BASE COUNT 104 a 196 c 198 g 154 t 2 others

ORIGIN

Query Match 95.0%; Score 19; DB 12; Length 654;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtggccgagagcagagaga 19
|||||
Db 192 GTGCCGAGCAGCAGAGAGA 174

RESULT 7 401 bp mRNA linear EST 21-Apr-1998
U80749/c

LOCUS U80749 Human fetal brain (R.L.Margolis) Homo sapiens cDNA, mRNA
DEFINITION sequence.

ACCESSION U80749
VERSION U80749.1 GI:2565071
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 401)
TITLE Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S.,
JOURNAL Breschel, R.S., Stine, O.C., Callahan, C., McInnis, M.G., and Ross, C.A.
MEDLINE CDNs with long CAG trinucleotide repeats from human brain
COMMENT Hum. Genet. 100 (1), 114-122 (1997)
97369492
CONTACT: Russell L. Margolis
Johns Hopkins University School of Medicine
720 Rutland Avenue, Baltimore, MD 21205-2196, USA
Tel: 410-614-0012
Fax: 410-614-0013
Email: rmargoli@welchlink.welch.jhu.edu.
FEATURES
source
1. 401
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6"
/clone_lib="Human fetal brain (R.L.Margolis)"
/tissue-type="brain"
/dev_stage="fetal"
BASE COUNT 85 a 118 c 120 g 68 t 10 others
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 401;
Best Local Similarity 95.0%; Pred. No. 2.7e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gtggcggagcagcagagac 20
||| ||||| ||||| ||||| |||||
DB 207 GTGCCCGAGCAGCAGCAGAC 188
RESULT 8 482 bp mRNA linear EST 29-FEB-2000
AL039573 DKF2434D1311.F1 434 (synonym: htes3) Homo sapiens cDNA clone
LOCUS DKF2434D1311 5', mRNA sequence.
DEFINITION
ACCESSION AL039573
VERSION AL039573
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 482)
Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann
S.
EST (Duesterhoeft, et al.)
TITLE Unpublished (1999)
JOURNAL Contact: Duesterhoeft A
COMMENT MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ). Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No 5' sequence available.
This clone (DKF2434D1311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1. 482
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="DKF2434D1311"
/clone_lib="434 (synonym: htes3)"
/tissue-type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPort1; Site_1: NotI; Site_2: SalI"
BASE COUNT 49 a 218 c 145 g 70 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 9; Length 482;
Best Local Similarity 95.0%; Pred. No. 2.8e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gtggcggagcagcagagac 20
||| ||||| ||||| ||||| |||||
DB 480 GTGCCCGAGCAGCAGCAGAC 461
RESULT 9 500 bp mRNA linear EST 05-SEP-2001
B1547486/c 603191091F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262335 5',
LOCUS mRNA sequence.
DEFINITION
ACCESSION B1547486
VERSION B1547486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 500)
TITLE NIH-MGC http://mhc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-riemail.nih.gov
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihaki
Toshiyuki and Piero Garinca (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11661 row: e column: 24
High quality sequence stop: 485.
Location/Qualifiers
1. 500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262335"
/clone_lib="NIH_MGC_95"
/tissue-type="hippocampus"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtggag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 57 a 222 c 150 g 71 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 500;
Best Local Similarity 95.0%; Pred. No. 2.8e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 gtggcggagcagcagagac 20

```

DB      489 GGGGCCGAGCAGCAGGAGAC 470
|||||
RESULT 10
LOCUS   A1513242
DEFINITION GH13253.3prime GH Drosophila melanogaster head POT2 Drosophila
          melanogaster cDNA clone GH13253 3prime, mRNA sequence.
ACCESSION A1513242
VERSION   A1513242.1
KEYWORDS EST.
SOURCE    fruit fly.
          Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 567)
AUTHORS  Harvey,D., Brokslein,P., Hong,L., Evans-Holm,M., Su.,, Tsang,G.,
          Lewis,S. and Rubin,G.M.
          BDGP/HMI Drosophila EST Project
          Unpublished (2001)
          Other_ESTs: GH13253.5prime
          Contact: Stapleton, M.
          BDGP
          Lawrence Berkeley National Lab
          One Cyclotron Rd, Berkeley, CA 94720, USA
          Fax: 510 486 6798
          Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
          Based upon the presence of a XhoI site followed by a run of 14 or
          more T residues at the beginning of the sequence, this clone was
          polyadenylated. The resulting Poly-T sequence has been removed. hit
          genomic sequence AC005440
          Plate: 132 row: E column: 5
          High quality sequence stop: 280.
          Location/Qualifiers
            1..567
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_lib="GH Drosophila melanogaster head POT2"
              /sex="male and female"
              /dev_stage="adult"
              /lab_host="DH5 - alpha"
              /note="Organ: head; Vector: POT2; Site.1: EcoRI; Site.2:
              XhoI; Sized fractionated cDNAs were directly ligated into
              POT2. Plasmid cDNA library."
BASE COUNT      153 a      116 c      124 g      174 t
ORIGIN
1 gtcgcgagacgagacgac 20
Db      488 GAGCCGAGCAGCAGGAGAC 469
|||||
RESULT 11
LOCUS   BE585103
DEFINITION 632 bp mRNA linear EST 16-AUG-2000
ACCESSION BE585103
VERSION   BE585103.1
KEYWORDS EST.
SOURCE    Phytophthora sojae.
          Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
          Phytophthora.
REFERENCE 1 (bases 1 to 632)
AUTHORS  Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.

```

```

TITLE      Comparative analysis of expressed sequences in phytophthora sojae
JOURNAL    Plant Physiol. 123 (1), 243-254 (2000)
MEDLINE    20267956
COMMENT    Contact: Gijzen M
          Agriculture and Agri-Food Canada
          1391 Sandford Street, London, Ontario, Canada N5V 4T3
          Tel: 519 457 1470
          Fax: 519 457 3997
          Email: gijzenm@em.agr.ca.
FEATURES   location/Qualifiers
SOURCE      1..632
            /organism="phytophthora sojae"
            /strain="race 2, strain P6497"
            /db_xref="taxon:67593"
            /clone_lib="PsojaeZO"
            /dev_stage="zoospores"
            /lab_host="E. coli strain XL0LR"
            /note="Vector: pBK-CMV; Site.1: EcoRI; Site.2: XhoI; This
            cDNA library was constructed from polyA+ enriched mRNA
            from zoospores grown in liquid medium. Zoospores were
            released into water and collected by centrifugation at
            2,000g; zoospore-bearing sporangia were induced to
            develop on 5 to 7 d old mycelium colonies grown on V8
            agar by repeated flooding with water. Complementary DNA
            was synthesized from mRNA using an XhoI-poly(TT)
            linker-primer. EcoRI adapters were ligated to the
            blunt-ended cDNA fragments and the products were digested
            with XhoI for directional cloning into lambda ZAP Express
            vector. This lambda library was amplified once using E.
            coli host strain XL1 Blue MRF+. Inserts were then
            subcloned by mass excision using Exsist1 helper phage
            for conversion into phagemid vector pBK-CMV in E. coli
            host strain XL0LR. Sequenced using T3 primer: 5' ATT AAC
            CCT CAC TAA AGG GA 3'."
BASE COUNT      131 a      168 c      207 g      124 t      2 others
ORIGIN
1 gtcgcgagacgagacgac 20
Db      258 GTGAGCAGCAGCAGGAGAC 277
|||||
RESULT 12
LOCUS   AM845240
DEFINITION 660 bp mRNA linear EST 19-MAY-2000
ACCESSION AM845240
VERSION   AM845240.1
KEYWORDS EST.
SOURCE    human.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 660)
AUTHORS  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
          Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
          Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
          Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          JOURNAL    MEDLINE
          COMMENT

```

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=et2-QV0-CT0018-011199-042-a06&rl=1999-11-01&rl=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 659.
Location/Qualifiers

FEATURES

source

1. 660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0018"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 114 a 253 c 187 g 106 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 660;
Best local Similarity 95.0%; Pred. No. 3e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtggcggagagcaggagac 20
|||||
Db 210 GTGCCGAGCAGCAGCAGC 191

RESULT 13
AUI24593/c 854 bp mRNA linear EST 23-OCT-2000
LOCUS AUI24593 NT2RM4 Homo sapiens cDNA clone NT2RM4000251 5', mRNA
DEFINITION
sequence.
ACCESSION AUI24593
VERSION AUI24593.1 GI:10949309
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 854)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- a 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
Location/Qualifiers

FEATURES

source

1. 854
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4000251"

/clone_lib="NT2RM4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"
BASE COUNT 194 a 303 c 233 g 123 t 1 others
ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 854;
Best local Similarity 95.0%; Pred. No. 3.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtggcggagagcaggagac 20
|||||
Db 52 GTGCCGAGCAGCAGCAGC 33

RESULT 14
BG387406/c 874 bp mRNA linear EST 12-MAR-2001
LOCUS 602456172P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4584448 5',
DEFINITION
mRNA sequence.
ACCESSION BG387406
VERSION BG387406.1 GI:13280852
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 874)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1309 row: p column: 17
High quality sequence stop: 531.
Location/Qualifiers

FEATURES

source

1. 874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4584448"
/clone_lib="NIH_MGC_15"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 225 a 287 c 253 g 109 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 874;
Best local Similarity 95.0%; Pred. No. 3.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gtggcggagagcaggagac 20
|||||
Db 200 GTGCCGAGCAGCAGCAGC 181

```

RESULT 15
BF204258          910 bp  mRNA  linear  EST 06-NOV-2000
LOCUS             601867720F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4110365 5',
DEFINITION        mRNA sequence.
ACCESSION         BF204258
VERSION           BF204258.1 GI:11097844
KEYWORDS
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 910)
AUTHORS           NIH-MGC http://mgc.nci.nih.gov/.
TITLE             National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL           Unpublished (1999)
COMMENT           Contact: Robert Strausberg, Ph.D.
                   Email: cgapbs-remail.nih.gov
                   Tissue Procurement: ATCC
                   cDNA Library Preparation: Ling Hong/Rubin Laboratory
                   cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                   DNA Sequencing by: Incyte Genomics, Inc.
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
                   Plate: LHC996 row: 9 column: 06
                   High quality sequence stop: 606.
FEATURES
   source
       1..910
           location/Qualifiers
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4110365"
               /clone_lib="NIH_MGC_17"
               /tissue_type="rhabdomyosarcoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                   Site_2: XhoI; cDNA made by oligo-dT priming.
                   Directionally cloned into EcoRI/XhoI sites using the
                   following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                   for average insert size 1.8kb. Library constructed by
                   Ling Hong in the laboratory of Gerald M. Rubin (University
                   of California, Berkeley) using ZAP-cDNA synthesis kit
                   (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      197 a      244 c      344 g      123 t      2 others
ORIGIN
Query Match      92.0%; Score 18.4; DB 10; Length 910;
Best Local Similarity 95.0%; Pred. No. 3.1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gtggcgcgagcagcagcagc 20
    |||||
Db 519 gtgctcgcgagcagcagcagc 538

```

Search completed: August 14, 2002, 21:04:06
 Job time: 10994 sec

Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
location/Qualifiers
source 1..264
/organism="Papio hamadryas"
/db_xref="taxon:9557"
<1..>264
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 25 a 130 c 78 g 31 t
ORIGIN

Query Match 100.0%; Score 31; DB 9; Length 264;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcggcctcccccctcgtcgtcc 31
|||||
3 CTCGGCGGCGCTCCCGCCCTTCGTCGCC 33

RESULT 2
AF330033 322 bp DNA linear PRI 08-NOV-2001
LOCUS Macaca radiata SCA2 gene, partial sequence.
DEFINITION AF330033
ACCESSION AF330033
VERSION AF330033.1 GI:12382835
KEYWORDS
SOURCE bonnet macaque.
ORGANISM Macaca radiata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 322)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
JOURNAL 2 (bases 1 to 322)
Choudhry,S. and Brahmachari,S.K.
DEFINITION Direct Submission
ACCESSION Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
KEYWORDS location/Qualifiers
SOURCE 1..322
/organism="Macaca radiata"
/db_xref="taxon:9548"
<1..>322
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 32 a 155 c 95 g 40 t
ORIGIN

Query Match 100.0%; Score 31; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcggcctcccccctcgtcgtcc 31
|||||
26 CTCGGCGGCGCTCCCGCCCTTCGTCGCC 56

RESULT 3
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS Presbytis entellus SCA2 gene, partial sequence.
DEFINITION AF330030
ACCESSION AF330030
VERSION AF330030.1 GI:12382832

KEYWORDS
SOURCE Hanuman langur.
ORGANISM Presbytis entellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Colobinae; Presbytis.
REFERENCE 1 (bases 1 to 384)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
JOURNAL 2 (bases 1 to 384)
Choudhry,S. and Brahmachari,S.K.
DEFINITION Direct Submission
ACCESSION Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
KEYWORDS location/Qualifiers
SOURCE 1..384
/organism="Presbytis entellus"
/db_xref="taxon:9574"
<1..>384
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 46 a 178 c 109 g 51 t
ORIGIN

Query Match 100.0%; Score 31; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcggcctcccccctcgtcgtcc 31
|||||
3 CTCGGCGGCGCTCCCGCCCTTCGTCGCC 33

RESULT 4
AF330029 409 bp DNA linear PRI 08-NOV-2001
LOCUS Gorilla gorilla SCA2 gene, partial sequence.
DEFINITION AF330029
ACCESSION AF330029
VERSION AF330029.1 GI:12382831
KEYWORDS
SOURCE
ORGANISM gorilla.
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Gorilla.
REFERENCE 1 (bases 1 to 409)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
JOURNAL 2 (bases 1 to 409)
Choudhry,S. and Brahmachari,S.K.
DEFINITION Direct Submission
ACCESSION Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
KEYWORDS location/Qualifiers
SOURCE 1..409
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
<1..>409
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 35 a 196 c 120 g 58 t
ORIGIN


```

Query March 100.0%; Score 31; DB 2; Length 231756;
Best Local Similarity 100.0%; Pred. NO. 0.66;
Matches 31; Conservative 0; Mismatches 0; Indels 0;
Gaps 0.

Qy 1 ctgcgcggagcctccgcgccttcgtgcgc 31
      |||||
db 89335 ctgcgcggagcctccgcgccttcgtgcgc 89305

```

RESULT	6
ARI59544	
LOCUS	ARI59544 355 bp DNA
DEFINITION	Sequence 1 from patent US 6251589.
ACCESSION	ARI59544
VERSION	ARI59544.1 GI:16222225
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
	Unclassified.
PAT	17-OCT-2001
linear	

REFERENCE	AUTHORS	TITLE	JOURNAL FEATURES	source	BASE COUNT	ORIGIN
1 (bases 1 to 355)	Tsuji, S. and Sampei, K.	Method for diagnosing spinocerebellar ataxia type 2 and primers therefor	Patent: US 6251589-A 1 26-JUN-2001;	1. .355 Location/Qualifiers	20 a	176 c 102 g 55 t 2 others
				"/organism="unknown"		

Query Match	96.8%	Score 30;	DB 6;	Length 355;
Best Local Similarity	100.0%	Pred. No.	5.6;	
Matches 30;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 ctcgagcggacccctcccgccccttgatgc 30
 |||||
Db 149 ctcggcgccccctccccccttccgtgcgc 178

RESULT	7				
AR159558					
LOCUS	AR159558	572 bp	DNA		
DEFINITION	Sequence	18 from patent US 6251589.		linear	PAT 17-OCT-2001
ACCESSION	AR159558				
VERSION	AR159558.1	GI:16222251			
KEYWORDS					
SOURCE	Unknown.				

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	SOURCE	BASE COUNT	ORIGIN
1 (bases 1 to 572)	Tsujii, S. and Sanpei, K.	Method for diagnosing spinocerebellar ataxia type 2 and primers therefor	Patent: US 6251589-A 18 26-JUN-2001;	Location/Qualifiers	1..572	34 a	277 c 174 g 85 t 2 others
					"/organism="unknown"		

Query Match	96.8%	Score 30;	DB 6;	Length 572;
Best Local Similarity	100.0%;	Pred. No. 5.1;		
Matches 30; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	ctcgagcgagcctcccgcccttcgtcgc	30
Db	149	ctcgagcgagcctcccgcccttcgtcgc	178

RESULT	8
ARI59546	

LOCUS	AR159546	623 bp	DNA	linear	PAT 17-OCT-2001
DEFINITION	Sequence 5 from patent US 6251589.				
ACCESSION	AR159546				
VERSION	AR159546.1	GI:16222229			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 623)				
TITLE	Tsujii, S. and Sempel, K.				
JOURNAL	Method for diagnosing spinocerebellar ataxia type 2 and primers				
FEATURES	therefor Patent: US 6251589-A 5 26-JUN-2001;				
source	Location/Qualifiers				
	1..623				
BASE COUNT	55 a 292 c 189 g 85 t			2 others	
ORIGIN	/organism="unknown"				

Query Match	96.8%	Score 30	DB 6	Length 623
Best Local Similarity	100.0%	Pred. No. 5		
Matches 30; conservative 0; Mismatches				
Indels	0			
Gaps	0			
Qy	1	ctcgcggagcctcccccgcctctgctgc	30	
Db	149	ctcgcggagcctcccccgcctctgctgc	178	

RESULT	9		
HDANSCA2			
LOCUS	HDANSCA2	4163 bp	
DEFINITION	H.sapiens mRNA for SCA2 protein.	linear	PRI 09-JAN-1997
ACCESSION	Y08262		
VERSION	Y08262.1		
KEYWORDS	SCA2 gene.		
SOURCE	human.		
ORGANISM	Homo sapiens		

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE
1 (bases 1 to 4163)	Imbert,G., Saudou,F., Yvert,G., Devys,D., Trottier,Y., Garnier,J.M., Weber,C., Mandel,J.L., Cancel,G., Abbas,N., Duerr,A., Diderjean,O., Stevanin,G., Agid,Y. and Brice,A.	Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats	Nat. Genet. 14 (3), 285-291 (1996)	97051922
2 (bases 1 to 4163)				

NUMBER	24262 (1/9)
TITLE	Direct Submission
JOURNAL	Submitted (20-SEP-1996) G. Imbert, I.G.B.M.C., Department Of Genetics, B.P. 163, 67404 Illkirch Cedex, FRANCE
FEATURES	Location/Qualifiers
SOURCE	1. .4163

```

/organism="Homo sapiens"
/isolate="DAN patient"
/db_xref="taxon:9606"
/cell_line="lymphoblastoid"
/clone_idb="DAN"
/dev_stage="adult"
1..2747
gene
/feature="SCA2"
<1..2747
CDS

```

```

/gene="SCA2"
/codon_start=3
/protein_id="CA69588.1"
/db_xref="GI:1770390"
/db_xref="SPTREMBL:Q99493"
/translacion="GNGCGAARPGRSLRLGCGPRPFVYLLPLASPGAPPAAPTRA
SPICARASPRSGVSLARPAPGCPRPACPEYVGLTMSLRPQOOOOOOOOOOOOOOO
OOOOOOPPAANVRKPGSGGILASPPAASPSSSSVSSEATAPSVAAATGGGGGGR
LGRNRSKGLPOSTISFDGIVAMRWHLITTSVGSCKCEQVYNGGIIYGEVFTYGP

```


KCDLVDAHEKSTESSGPKREFIMESILFKCSDPVVVOFKMDSSAKRDAFTDSA
ISAKVNGHEKLEPMDAGELTANIELEJALENDVSNKWDNDNRIVEENYGVASTY
DSSLSTVPLERDNSEEFLEKREARANOJLAEISSAQYKARVALENDREERKTYA
VORNSSEREGSHINTREKXYIPGCRNREYVISMCGROKNSPMGCPGSGMSSTSH
TSDPNNSGDORVNGVWPVSPCPSPSPRYOGSPNLSPPRAATPPRPSRPP
SRPSRPSHPSPAGSPAPVSPMPKRMSESGPRMSPKORHRNHRVAGSGTSSGL
EYVHNPPSEATPPVARTSPSGGTSWSSVYGRSLSKORHRNHRVAGSGTSSGL
VLASPOAGITTEAVNAPITPAASPTPASPANRATVPSSEAKDSRLDQRONSPAGNK
ENIKPNETSPSPSKAENKGISPVVSEHRKQIDDLKKEKNDRIQSPSTSEMDOLNK
NRECKSHDLIKDKIEPSAKDSFTIENSSNCTSGKNSPISILSNTBKRGP
VTSQVOTSPACKOEKDKDAEOVKSTLNPRAKFNPRSPFOQPPSTPTS
PRQAPSPSPSVGHQOPTPVYTOPVCFAPNMYVPVPSGVQYDLCPSGKSTIIRVP

BASE COUNT 1136 a 1196 c 908 g 923 t

ORIGIN

Query Match 96.8%; Score 30; DB 9; Length 4163;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgagctcccccctcgcgc 30
|||||

Db 51 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 80

RESULT 10
LOCUS A62706 4200 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 7 from Patent WO9717445.
ACCESSION A62706
VERSION A62706.1 GI:3716590
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 4200)
AUTHORS Tori,L., Butz,Y., Trotter,Y., Mandel and Jean-Louis.
TITLE METHOD FOR TREATING NEURODEGENERATIVE DISEASES USING A 1C2 ANTIBODY
OR A FRAGMENT OR DERIVATIVE THEREOF, AND CORRESPONDING
PHARMACEUTICAL COMPOSITIONS
JOURNAL Patent: WO 9717445-A 7 15-MAY-1997;
COMMENT CENTRE NAT RECH SCIENT (FR)
FEATURES Other publication FR 2741088 19970516.
source Location/Qualifiers
1..4200
/organism="unidentified"
/db_xref="taxon:32644"
/clone="DAN1"

BASE COUNT 1152 a 1200 c 913 g 935 t

ORIGIN

Query Match 96.8%; Score 30; DB 6; Length 4200;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgagctcccccctcgcgc 30
|||||

Db 51 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 80

RESULT 11
LOCUS ARI53580 4481 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 18 from patent US 6235872.
ACCESSION ARI53580
VERSION ARI53580.1 GI:15121112
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4481)

AUTHORS Bredesen,D.E. and Rabilzadeh,S.
TITLE Prapaploctic peptides dependence polypeptides and methods of use
JOURNAL Patent: US 6235872-A 18 22-MAY-2001;
FEATURES Location/Qualifiers
source 1..4481
/organism="unknown"

BASE COUNT 1144 a 1380 c 1014 g 943 t

ORIGIN

Query Match 96.8%; Score 30; DB 6; Length 4481;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgagctcccccctcgcgc 30
|||||

Db 451 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 480

RESULT 12
LOCUS HSU70323 4481 bp mRNA linear PRI 20-NOV-1996
DEFINITION Human ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION U70323
VERSION U70323.1 GI:1679683
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4481)
AUTHORS Pulst,S.-M., Nechiporuk,A., Nechiporuk,T., Gispert,S., Chen,X.-N.,
Lopes-Cendes,L., Pearlman,S., Starkman,S., Orozco-Diaz,G.,
Lunkes,A., DeJong,P., Rouleau,G.A., Auburger,G., Korenberg,J.R.,
Figueroa,C. and Saba,S.
TITLE Moderate expansion of a normally diallelic trinucleotide repeat in
spinocerebellar ataxia type 2
JOURNAL Nature Genet. 14 (3), 269-276 (1996)
REFERENCE 2 (bases 1 to 4481)
AUTHORS Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1996) Medicine, Cedars-Sinai, 8700 Beverly Blvd.,
Los Angeles, CA 90048, USA
FEATURES Location/Qualifiers
1..4481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24.1"
1..4481
/gene="SCA2"
163..4101
/gene="SCA2"
/standard_name="spinocerebellar ataxia type 2"
/codon_start=1
/product="ataxin-2"
/protein_id="AAB19200.1"
/db_xref="GI:1679684"
/translation="MRSAARPAVATETSRFAARWPMGMSLORPARSRGGGG
AARGPYSAAPPDPGPPSPKSSQSSPPSADDFGNSNGGARFGSGRRLLGPPR
PVTYVLLPLASFGAPPAAPTRASPAGARSPRSGVSLARPAAGCPRRACEVYTPLT
MSLRPQ000000000000000000000000PPAPAAVNRKPGSGSLASPAAPSPSSSV
SSSSATAPSSVVAATSGGGRPGDLGRGNSKGLPQSTISFDGIYANMRVHLTVSG
SKCEVQVANGGIVYEGFTYSPKCDLVDAHEKSTESSGPKREFIMESILFKCSD
VYVQFKMDSSYAKRDAFTDSAISAKVNGHEKLEKDEPMDAGELTANIELEJALENDV
NGMDNDMEFRVMEYGVSTYDSSLSTVPLERDNSEEFLEKREARANOJLAEISS
AOYKARVLENDRESEERTYAVQKNSEREGHSTINENKITIPGQKREYVTSWG
RONSPPKOPGSGMSPTSHSDFNPNNSGDORVNGVWPVSPCPSPSPSRYSY
OSGPNLSPPRAATPPRPSRPSRPSRPSHPAGSPAPVSTMPKRMSESGPRMSP
KAORHRNHRVAGSGTSSGLIEFVSHNPSPAAIPVARTSPSGTSSVYGVART
SPKTRRPSRONSIGNTSPGVLASPOAGITTEAVNAPITPAASPTPASPANRAT
PSEAKDSRLDQRONSPAGNKENIKPNETSPSPSKAENKGISPVVSEHRKQIDDLK

FKNDRLQPSSTSESMDQLNKNREGEKSRDLINKIEPSAKDSFIENSSNCTSGSS
KPNPSISPSILSTNTEHRCGEVTSOGVQSSPACQOEKDEKEDKAAAEVRSSTLN
PNKRFNPNRSOPKPTSTTPSPRPQOQSSMVGHOOPVTYTOPCFEAFNNMYVPV
VSRGQVATSPQOFPNPLVQHPHYOSCPHYSPYISYIGNAMMAPTIAOGCLVS
PASYTOYGAHEQTAMACPKLYNKETSPSTFAISTOSLAQOYAHNPATLHPHP
QPSATPTCOQOOSQHGSHAPSPVQHQAALHLASPOQSAITYHAGLAPTPSM
TPASNTQSPQNSFPAAQOQVPTIHPSHVQPAVTNPMHAPHVQAHVQSGMVPSPHTAH
APMLMTQPPGPGQALAQSLQPIPTSTAHFPYMTHPSPVQAHHOOL"

BASE COUNT 1144 a 1380 c 1014 g 943 t

ORIGIN

Query Match 96.8%; Score 30; DB 9; Length 4481;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgagcccccgccttcgtcgc 30
|||||
DB 451 CTCGGCGGCGCTCCCGCCCTTCGTCGC 480

RESULT 13
AF330031 303 bp DNA linear PRI 08-NOV-2001
LOCUS AF330031
DEFINITION Macaca mulatta SCA2 gene, partial sequence.
ACCESSION AF330031
VERSION AF330031.1 GI:12382833
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.

REFERENCE 1 (bases 1 to 303)
Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
Brahmachari, S.K.

TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490

REFERENCE 2 (bases 1 to 303)
Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
1..303
Location/Qualifiers

gene
/organism="Macaca mulatta"
/db_xref="taxon:9544"
<1..>303
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 32 a 143 c 92 g 36 t

ORIGIN

Query Match 87.1%; Score 27; DB 9; Length 303;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gcgggcctccgcgccttcgtcgc 31
|||||
DB 1 GCGGGCTCCCGCCCTTCGTCGC 27

RESULT 14
AF330028 390 bp DNA linear PRI 08-NOV-2001
LOCUS AF330028
DEFINITION Pan troglodytes SCA2 gene, partial sequence.
ACCESSION AF330028
VERSION AF330028.1 GI:12382830

KEYWORDS
SOURCE
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

REFERENCE 1 (bases 1 to 390)
Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
Brahmachari, S.K.

TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490

REFERENCE 2 (bases 1 to 390)
Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
1..390
Location/Qualifiers

repeat_region
1..390
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/note="microsatellite"
/rpt_type=tandem
/rpt_unit=cag
<1..>390
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 48 a 183 c 110 g 49 t

ORIGIN

Query Match 87.1%; Score 27; DB 9; Length 390;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gcgggcctccgcgccttcgtcgc 31
|||||
DB 1 GCGGGCTCCCGCCCTTCGTCGC 27

RESULT 15
AF041472 4225 bp mRNA linear ROD 28-NOV-2001
LOCUS AF041472
DEFINITION Mus musculus ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION AF041472
VERSION AF041472.1 GI:3005019
KEYWORDS
SOURCE
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4225)
Nechiporuk, T.T., Huynh, D.P., Figueroa, K., Sabha, S., Nechiporuk, A.V.
and Pulst, S.M.
TITLE The mouse SCA2 gene: cDNA sequence, alternative splicing and
protein expression
JOURNAL Hum. Mol. Genet. 7 (8), 1301-1309 (1998)
MEDLINE 98334550
PUBMED 9668173

REFERENCE 2 (bases 1 to 4225)
Nechiporuk, T.T., Figueroa, K., Sabha, S., Nechiporuk, A.V. and
Pulst, S.M.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Medicine/Neurology, Cedars-Sinai Medical
Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA

FEATURES
source
1..4225
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="12q23.1"

gene 1. .4225
/gene="SCA2"
27. .3884
CDS /gene="SCA2"
/codon_start=1
/product="ataxin-2"
/protein_id="AAC09275.1"
/db_xref="GI:3005020"

/translation="MSSSTAAVORPAAGDPEPRRPAAGMARSLPRTARRGGRGAVA
YPSAGPPRPGAPRRGPRSPCASDCFCSSNGHAGRRPSRLICGPPRPVYLL
ALAPATPARACPPGVASPPRGVSSSARPAAGCPACPEYTGFLTMSLKPQPP
APATGRKPGGLSPGAPASAAVTSASVAPAPAVASSSAAAGGRRGLGRNS
SKGLPQITISFDGIYANVMVHILTSVSGKCEVYKNGIYEGVEKITYSPKDLVLD
AAHEKTESSSGPREIMEYLFKCSDFVYQFKDTSYARDAFTDSALSAKVG
EHKEDLEPDAGELTASPELELENDVNGMDNDMFYNEENYGVSTYDSLSYT
VLEPRNSEEFLLKREARANOIAEEIESSAOYKARVALENDRSEEEKYTAOYRNCSDR
EGHGPTKDNKTIIPQQRNREYLSMGSROSSPRMGQPGGSMPSRAASHTSDPNNA
GSDQRYVNGVMPSPSPSHSSRPSPRYOSGPNSLPPRAATHTRPSPSPSRPSP
HPSAHGSPAPVSTMPKRMSEGGPRMSPRAQRHPNHRVYAGKSMGSLFVSHNP
SEAAAPPVARTSPAGGTMSVSGVRLSPKTHRPSPRQSSIGNSPSGVLASPOAG
IIPAEAVMPVPASPTPASPNRALTPTSEAKDSRLQDORONSPAGSKENYKASET
SPFSKADNKGMSPVYSEHRKQIDDLKKFKNDRRLQPSSTSEMDQLSKNREGKSR
DLIKDKTEASAKDSFIDSSSSSNCSTGSSKTNPSISPMLSNAEHKRGPEYTSQV
QTSPPACKQEKDDREKIDTTEOVKSTLNPNAKENPNSFSQPKSTPTSFRPQAQ
PSPSMVGHQOPAPVYTQVPCFAPNMMY PVPVSPGVQLYPIPTMPVNOAKTYRAGK
VPMPOORODOHOSSTMHPASAGPPIVATPPAYSTOYVAYSPOQFPNOPLVOHYPH
YOSORPHVYSPVIOGNARMMAPPAAHAPGLVSSSAQFGAHEQTHAMVACPLPYNKE
TSPSEYFAISTGSLAQOYAHNPALHPHTPHQPSATPTGGOQSOHGGSHAPSPVOH
HOKQAQALHLASPOOQSAIYHAGILAPTPSPMPASNTQSPSSPPAAOQTVTTIHS
HVOPATTPPHMAHVPOAHVQSGMVPSTPAHAPMLMTTQPPGPKAALQSAADPIP
VSTTAHPPTMTHPSVOAHHQQL"

BASE COUNT 1007 a 1324 c 1042 g 851 t 1 others
ORIGIN

Query Match 85.2% Score 26.4; DB 10; Length 4225;
Best Local Similarity 96.4% Pred. No. 44;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 tcggcgaggcctccggcccttcgtcgt 29
|||
Db 295 TCTGGGGCCTCCCGCCCTTCGTGCT 322

Search completed: August 14, 2002, 21:48:07
Job time: 13505 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 22:06:20 : Search time 906.46 Seconds
(Without alignments)
58.717 Million cell updates/sec

Title: US-09-707-919-3
Sequence: 1 ctccgcgcgcctcccccctcgtcgtcc 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	96.8	355	19 AAV17224	SCA2 gene fragment
2	30	96.8	516	19 AAV06551	SCA2 gene fragment
3	30	96.8	623	19 AAV17229	SCA2 gene fragment
4	30	96.8	4200	18 AAT78912	Spinocebellar at
5	30	96.8	4367	19 AAV30270	Gene causative of
6	30	96.8	4481	19 AAV06552	Human SCA2 cDNA in
7	30	96.8	4481	20 AA23428	Human SCA2 cDNA. H
8	21.4	69.0	328	22 AAL34906	Human musculoskele
9	20.6	66.5	427	19 AAV62163	HSV-2 strain SB5 C

C 10	20.6	66.5	726	22 AAK89967	Human digestive sy
C 11	20.6	66.5	1008	21 AAC56198	Eucalyptus grandis
C 12	20.6	66.5	3957	22 AAA09686	HSV-2 immediate ea
C 13	20.6	66.5	16812	19 AAV62175	HSV-2 strain SB5 C
C 14	20.6	66.5	154746	24 AAD25519	Human herpesvirus
C 15	20.6	66.5	154746	24 AAD25519	Human herpesvirus
C 16	20.4	65.8	3592	23 ABL13077	Drosophila melanog
C 17	20.4	65.8	3994	21 AAC76475	Human ORFX ORF2030
C 18	20.4	65.8	17500	23 ABL13076	Drosophila melanog
C 19	20.2	65.2	796	22 AAL36153	Human musculoskele
C 20	20.2	65.2	2930	22 AAL36154	Human musculoskele
C 21	20	64.5	1101	23 AAS66114	DNA encoding novel
C 22	20	64.5	1101	23 AAS71835	DNA encoding novel
C 23	19.8	63.9	1214	22 AAS26402	Human cDNA encodin
C 24	19.8	63.9	1434	21 AAC56333	Pinus radiata tran
C 25	19.8	63.9	1730	22 AAS25963	Human cDNA encodin
C 26	19.8	63.9	1821	21 AAC55846	Mitomycin biosynth
C 27	19.8	63.9	1945	22 AAK52723	Human polynucleoti
C 28	19.8	63.9	7419	23 AAS51427	Pseudomonas derugi
C 29	19.8	63.9	18034	21 AAC55841	Complete Mitomycin
C 30	19.8	63.9	133719	21 AAC64754	Macaca mulatta rha
C 31	19.8	63.9	349980	22 AAH41225	Pyrococcus abyssi
C 32	19.4	62.6	1139	9 AAN80100	Endo-beta-N-acetyl
C 33	19.4	62.6	1200	15 AAO54355	Rat post-synaptic
C 34	19	61.3	413	22 AAS45281	CDNA encoding nove
C 35	19	61.3	500	21 AAB65494	Porcine BAC-PiGF2
C 36	19	61.3	780	22 AAO6798	Human CDNA clone (
C 37	19	61.3	856	21 AAC75808	Human ORFX ORF1363
C 38	19	61.3	950	21 AAC56025	Eucalyptus grandis
C 39	19	61.3	1200	21 AAB65504	Porcine BAC-PiGF2
C 40	19	61.3	1311	22 AAS45093	CDNA encoding nove
C 41	19	61.3	3248	22 AAH14518	Human CDNA sequenc
C 42	19	61.3	3533	22 AAI71765	Human cancer-inhib
C 43	19	61.3	5276	20 AAX87397	Human WART2 cDNA.
C 44	19	61.3	5486	21 AAS59129	DNA encoding a tum
C 45	19	61.3	5486	21 AAS59130	DNA encoding a tum

ALIGNMENTS

RESULT 1	
AAV17224	AAV17224 standard; DNA; 355 BP.
ID	AAV17224
XX	AAV17224:
AC	29-JUN-1998 (first entry)
XX	
DE	SCA2 gene fragment.
XX	
KW	SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX	
OS	Synthetic.
XX	
FN	Key
FT	Location/Qualifiers
FT	CDS 341..355
FT	/*tag="a
FT	/note="SCA2 protein fragment"
XX	
PN	W09803679-A1.
XX	
PD	29-JAN-1998.
XX	
PF	18-JUL-1996; 96WO-JP01999.
XX	
PR	18-JUL-1996; 96WO-JP01999.
XX	
PA	(SRLS-) SRL INC.
XX	
PI	Samuel K, Tsuji S;
XX	
DR	WPI: 1998-120796/11.

DR	P-PSDB; AAM41370.
XX	
PT	Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT	number of CAG repeat units
XX	
PS	Claim 1: Page 10; 23pp; Japanese.
XX	
CC	This sequence represents a fragment of the SCA2 gene. It can be used in
CC	the method of the invention for diagnosing spinocerebellar ataxis type
CC	II, by performing PCR on the test DNA using two primers hybridising to
CC	parts of the SCA2 gene sequence, and determining the number of CAG
CC	repeats in the amplified products. The method provides an easy means for
CC	the diagnosis of spinocerebellar ataxis type II.
XX	
SO	Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;
Query Match	96.8%; Score 30; DB 19; Length 355;
Best Local Similarity	100.0%; Pred. No. 0.089;
Matches 30: Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ctcggcgggcgcccccgccttcgtcgc 30
Db	149 ctcggcgggcgcccccgccttcgtcgc 178
RESULT 2	
AAV06551	
ID	AAV06551 standard; DNA; 516 BP.
XX	
AC	AAV06551;
XX	
DT	06-JUL-1998 (first entry)
XX	
DE	SCA2 gene fragment including CAG repeat region.
XX	
SCA2 gene: spinocerebellar ataxia-2; ataxin-2; human;	
KW	diagnosis; olivoponto-cerebellar atrophy; ss; ds.
KW	
XX	
OS	Homo sapiens.
XX	
XX	
FH	Key
FH	primer_bind
FT	Location/Qualifiers
FT	/*tag= a complement (241..257)
FT	/note= "primer SCA2-A binding site"
FT	349..366
FT	/*tag= b
FT	/note= "primer SCA2-B binding site"
FT	499..500
FT	/*tag= c
FT	/note= "predicted splice site"
FT	267..332
FT	/*tag= d
FT	/note= "CAG repeat region"
FT	267..269
FT	/*tag= e
FT	/note= "CAG repeat"
FT	270..272
FT	/*tag= f
FT	/note= "CAG repeat"
FT	273..275
FT	/*tag= g
FT	/note= "CAG repeat"
FT	276..278
FT	/*tag= h
FT	/note= "CAG repeat"
FT	279..281
FT	/*tag= i
FT	/note= "CAG repeat"
FT	282..284
FT	/*tag= j
FT	/note= "CAG repeat"
FT	285..287

FT	/*tag=	k	
FT	/note=	"CAG repeat"	
FT	repeat_unit	291..293	
FT	/*tag=	1	
FT	/note=	"CAG repeat"	
FT	repeat_unit	294..296	
FT	/*tag=	m	
FT	/note=	"CAG repeat"	
FT	repeat_unit	297..299	
FT	/*tag=	n	
FT	/note=	"CAG repeat"	
FT	repeat_unit	300..302	
FT	/*tag=	o	
FT	/note=	"CAG repeat"	
FT	repeat_unit	306..308	
FT	/*tag=	p	
FT	/note=	"CAG repeat"	
FT	repeat_unit	309..311	
FT	/*tag=	q	
FT	/note=	"CAG repeat"	
FT	repeat_unit	312..314	
FT	/*tag=	r	
FT	/note=	"CAG repeat"	
FT	repeat_unit	315..317	
FT	/*tag=	s	
FT	/note=	"CAG repeat"	
FT	repeat_unit	318..320	
FT	/*tag=	t	
FT	/note=	"CAG repeat"	
FT	repeat_unit	321..323	
FT	/*tag=	u	
FT	/note=	"CAG repeat"	
FT	repeat_unit	324..326	
FT	/*tag=	v	
FT	/note=	"CAG repeat"	
FT	repeat_unit	327..329	
FT	/*tag=	w	
FT	/note=	"CAG repeat"	
FT	repeat_unit	330..332	
FT	/*tag=	x	
FT	/note=	"CAG repeat"	
XX			
PN	W09742314-A1.		
XX			
PD	13-NOV-1997		
XX			
PF	08-MAY-1997;	97WO-US07725.	
XX			
PR	08-OCT-1996;	96US-0727084.	
PR	08-MAY-1996;	96US-0017388.	
PR	19-JUL-1996;	96US-0022207.	
XX			
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.		
XX			
PI	Pulst S;		
XX			
DR	WPI: 1998-086523/08.		
XX			
PT	Nucleic acids encoding human and mouse ataxin 2 - a product of the		
PT	spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of the		
PT	ataxia type 2		
XX			
PS	Example 2; Page 51-52; 98pp; English.		
XX			
CC	This genomic DNA in plasmid PL65122B includes a CAG repeat region		
CC	from the novel human SCA2 gene (see AA06552). It was identified		
CC	following the construction of a bacterial artificial chromosome		
CC	contig and a pl artificial chromosome of the spinocerebellar		
CC	ataxia 2 (SCA2) gene region and the identification of the SCA2		
CC	gene from this contiguous map unit using a technique that screens		
CC	for the presence of DNA trinucleotide repeats. The SCA2 locus is		
CC	at 12q24.1. Ataxia type 2 can be diagnosed by detecting a genomic		
CC	or transcribed mRNA sequence in an individual having an expanded		

CC CAG repeat at a location corresponding to the CAG repeat region of
 CC the SCA2 gene. The presence of at least 13 CAG repeats above the
 CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
 CC primers (see AAT9640-41) amplifying at least this region are used
 CC for diagnosis. Also claimed are full-length ataxin-2 cDNAs for
 CC human and mouse (see AAV06552-53). Kits for detecting mutations at
 CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
 CC useful for studying the physiological roles of SCA2 polypeptide
 CC (ataxin-2, see AAM33807-08) and its effect upon behaviour.

CC
 XX
 SQ Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;

Query Match 96.8%; Score 30; DB 19; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.085;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctgcgggggctccccccttcgctgc 30
 ||||||||||||||||||||||||||||
 Db 60 ctgcgggggctccccccttcgctgc 89

RESULT 3
 AAV17229
 ID AAV17229 standard; DNA; 623 BP.

XX
 AC AAV17229;

DT 29-JUN-1998 (first entry)

XX SCA2 gene fragment.

KW SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

XX Synthetic.

FT Key Location/Qualifiers

FT CDS 341..583

FT /tag= a /note= "SCA2 protein fragment, no stop codon given"

FT WO9803679-A1.

XX 29-JAN-1998.

XX 18-JUL-1996; 96WO-JP01999.

XX 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sanpei K, Tsuji S;

XX WPI; 1998-120796/11.

XX DR P-PSDB; AAW41372.

XX PT Diagnosing spinocerebellar ataxis type II - by PCR and determining

XX PS Example 1; Page 11-12; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in
 CC the method of the invention for diagnosing spinocerebellar ataxis type
 CC II, by performing PCR on the test DNA using two primers hybridising to
 CC parts of the SCA2 gene sequence, and determining the number of CAG
 CC repeats in the amplified products. The method provides an easy means for
 CC the diagnosis of spinocerebellar ataxis type II.

XX Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 96.8%; Score 30; DB 19; Length 623;
 Best Local Similarity 100.0%; Pred. No. 0.085;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctgcgggggctccccccttcgctgc 30
 ||||||||||||||||||||||||||||
 Db 149 ctgcgggggctccccccttcgctgc 178

RESULT 4
 AAT78912
 ID AAT78912 standard; cDNA; 4200 BP.

XX AAT78912;

DT 09-FEB-1998 (first entry)

XX Spinocerebellar ataxia gene SCA2.

KW Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;

KW repeat region; affinity; RPA binding protein; Kennedy disease;

KW transcription initiation factor; lymphoblastic cell line; schizophrenia;

KW Huntington's disease; dominant autosomal spinocerebellar ataxia;

KW X-linked spino-bulbar muscular atrophy; familial spastic paraplegia;

KW dentatorubral-pallidolusial atrophy; bipolar affective disorder;

KW manic depressive psychosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 3..2747

FT /tag= a

FT /product= SCA2 protein

FT /note= "this CDS contains a putative translational start

FT codon for the SCA2 protein at positions 243-245"

FT CDS 2594..3640

FT /tag= b

FT /note= "this second open reading frame may be derived

FT by a frameshift or by alternative splicing"

FT CDS 3..242

FT /tag= c

FT /note= "putative open reading frame which is in frame

FT with the putative translational start site of

FT the SCA2 open reading frame"

FT misc-signal 239..245

FT /tag= d

FT /note= "putative Kozak consensus signal"

FT repeat-region 258..323

FT /tag= e

FT /note= "encodes polyglutamine repeat region; contains

FT repeats of CAG with 2 CAA codons interspersed"

FT repeat-unit 258..260

FT /tag= f

FT /note= "CAG repeats"

FT misc-feature 1..3986

FT /tag= g

FT /note= "sequence contained in DAN1 clone"

FT misc-feature 3987..4200

FT /tag= h

FT /note= "derived from the EST's AAH92640, AAN90240 and

FT AAZ13574 from dbEST database"

FT misc-feature 4023..4029

FT /tag= i

FT /note= "region which differs in length between the

FT sequences of the EST clones AAH92640, AAN90240

FT and AAZ13574"

XX WO9717445-A1.

XX 15-MAY-1997.

XX 08-NOV-1996; 96WO-FR01773.

XX 10-NOV-1995; 95FR-0013576.

PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX Lutz Y, Mandel J, Tora L, Trotter Y;
 XX
 DR WPI: 1997-281034/25.
 DR P-PSDB: AAM24800, AAM24801.
 XX
 PT Antibody 1C2 used for treating or preventing neuro-degenerative
 PT diseases - associated with proteins containing long poly:glutamine
 PT repeats, e.g. Huntington's disease
 XX
 PS Claim 21, Page 45-47; 69pp; French.
 XX
 CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
 CC treatment of neurodegenerative diseases associated with the presence
 CC of polyglutamine repeat regions. This Mab is already known for its
 CC affinity to the TATA binding protein (TBP) transcription initiation
 CC factor, especially at the amino acid sequence LEEQGRQ00000 found at
 CC the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
 CC for polyglutamine repeats with a proportional affinity to the number
 CC of glutamine repeats. This affinity has been used to identify genes
 CC encoding proteins containing long polyglutamine repeats which are
 CC implicated in neurodegenerative diseases. A screen of an expression
 CC library, generated from a lymphoblastic cell line from a patient
 CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
 CC new sequences (AA78906-T78911) encoding polyglutamine repeats. Mab 1C2
 CC also isolated the complete SCA2 gene in clone DNM1 (sequence presented
 CC here). The sequence appears to contain 2 open reading frames (ORF) the
 CC second of which may be generated by an frameshift slippage or by an
 CC alternative splicing event. The first ORF also encodes a 22 amino acid
 CC polyglutamine repeat region near the N-terminus of the protein. Normal
 CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
 CC interspersed whereas the mutant sequence from patients with SCA
 CC contains at least 30, preferably 37-50 CAG repeats.
 CC Mab 1C2, active fragment of it or nucleic acids encoding it are
 CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
 CC X-linked spinobulbar muscular atrophy (Kennedy disease),
 CC dentatorubral-pallidoluysal atrophy, dominant autosomal spinocerebellar
 CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
 CC depressive psychoses and schizophrenia.
 XX
 SQ Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

Query Match 96.8%; Score 30; DB 18; Length 4200;
 Best Local Similarity 100.0%; Pred. No. 0.073;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcgagggcctcccccctcgtcgtc 30
 ||||||||||||||||||||||||||||
 DB 51 ctgcgagggcctcccccctcgtcgtc 80

RESULT 5
 AAV30270
 ID AAV30270 standard; DNA; 4367 BP.
 XX
 AC AAV30270;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
 XX
 KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
 KW CAG repeat; neurodegenerative disease; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 49..3990
 FT /tag= a
 FT /product= "Spinocerebellar ataxia type 2 associated

FT repeat_region 544..612 protein"
 FT /tag= b
 FT /note= "normal CAG repeat region; this is increased in
 FT repeat_unlt 544..546 patients with SCA2"
 FT /tag= c
 XX
 PN W09818920-A1.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-JP03946.
 XX
 PR 30-OCT-1996; 96JP-0304059.
 XX
 PA (SRLS-) SRL INC.
 XX
 PI Sanpei K, Tsuji S;
 DR WPI: 1998-272215/24.
 DR P-PSDB: AAM60213.
 XX
 PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
 PT - contain increased number of CAG repeat region compared to normal
 PT gene
 XX
 PS Claim 1; Pages 13-22; 38pp; Japanese.
 XX
 CC This represents the sequence of a gene causative of spinocerebellar
 CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
 CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
 CC expression product produces a polyglutamine sequence from Gln-166 to
 CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
 CC patients this number is increased to 35-100. Peptides encoded by nucleic
 CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
 CC gene, antibodies recognising the peptides and antisense nucleic acids
 CC hybridising with the nucleic acid fragments can be used for the
 CC investigation and diagnosis of SCA2. They can also be used for the
 CC treatment of SCA2 by antisense therapy or gene therapy.
 XX
 SQ Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

Query Match 96.8%; Score 30; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred. No. 0.073;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgcgagggcctcccccctcgtcgtc 30
 ||||||||||||||||||||||||||||
 DB 337 ctgcgagggcctcccccctcgtcgtc 366

RESULT 6
 AAV06552
 ID AAV06552 standard; CDNA; 4481 BP.
 XX
 AC AAV06552;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human SCA2 CDNA including CAG repeat region.
 XX
 KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 164..4101
 FT /tag= a
 FT primer_bind complement (631..648)
 FT /tag= b

```

FT      primer_bind      /note="primer SCA2-A binding site"
FT      740..757
FT      /tag= c
FT      /note="primer SCA2-B binding site"
FT      1070..1091
FT      /tag= d
FT      /note="primer SCA2-14B binding site"
FT      899..900
FT      /tag= e
FT      /note="predicted splice site"
FT      658..723
FT      /tag= f
FT      /note="CAG repeat region"
FT      658..660
FT      /tag= g
FT      /note="CAG repeat"
FT      661..663
FT      /tag= h
FT      /note="CAG repeat"
FT      664..666
FT      /tag= i
FT      /note="CAG repeat"
FT      667..669
FT      /tag= j
FT      /note="CAG repeat"
FT      670..672
FT      /tag= k
FT      /note="CAG repeat"
FT      673..675
FT      /tag= l
FT      /note="CAG repeat"
FT      676..678
FT      /tag= m
FT      /note="CAG repeat"
FT      679..681
FT      /tag= n
FT      /note="CAG repeat"
FT      685..687
FT      /tag= o
FT      /note="CAG repeat"
FT      688..690
FT      /tag= p
FT      /note="CAG repeat"
FT      691..693
FT      /tag= q
FT      /note="CAG repeat"
FT      694..696
FT      /tag= r
FT      /note="CAG repeat"
FT      700..702
FT      /tag= s
FT      /note="CAG repeat"
FT      703..705
FT      /tag= t
FT      /note="CAG repeat"
FT      706..708
FT      /tag= u
FT      /note="CAG repeat"
FT      709..711
FT      /tag= v
FT      /note="CAG repeat"
FT      712..714
FT      /tag= v
FT      /note="CAG repeat"
FT      715..717
FT      /tag= x
FT      /note="CAG repeat"
FT      718..720
FT      /tag= y
FT      /note="CAG repeat"
FT      721..723
FT      /tag= z
FT      /note="CAG repeat"

```

```

XX      MO9742314-A1.
PN      13-NOV-1997.
PD      08-MAY-1997; 97MO-US07725.
XX      08-OCT-1996; 96US-0727084.
PR      08-MAY-1996; 96US-0017388.
PR      19-JUL-1996; 96US-0022207.
XX      (CEDA-) CEDARS SINAI MEDICAL CENT.
XX      Pulst S;
XX      MPI; 1998-086523/08.
DR      P-PSDB; AAW33807.
XX      Nucleic acids encoding human and mouse ataxin 2 - a product of the
PT      spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
PT      ataxia type 2
XX      Claim 6; Page 52-58; 98pp; English.
XX      This cDNA sequence corresponds to a novel SCA2 gene encoding a human
CC      spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
CC      (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
CC      human frontal cortex cDNA library in lambda ZapII were screened
CC      with probes obtained by PCR amplification of plasmid AAP651228 (see
CC      AAW06551). PCR products were used to screen the human adult frontal
CC      cortex library, and 5' clones were obtained by RT-PCR of placental
CC      mRNAs. Overlapping clones was used to generate the composite 4481
CC      bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
CC      or transcribed mRNA sequence in an individual having an expanded
CC      CAG repeat at a location corresponding to the CAG repeat region of
CC      the SCA2 gene. The presence of at least 13 CAG repeats above the
CC      normal level (22, occasionally 23, repeats) is indicative of SCA2.
CC      Primers (see AAT969640-41) amplifying at least this region are used
CC      for diagnosis. Also claimed are kits for detecting mutations at
CC      the SCA2 locus, antisense oligonucleotides, and transgenic animals
CC      useful for studying the physiological roles of ataxin-2 and its
CC      effect upon behaviour.
XX      SO      Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match      96.8%; Score 30; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ctggcgagcctcccgcccttcgtcgc 30
      |||
DB      451 ctggcgagcctcccgcccttcgtcgc 480

RESULT      7
AAZ3428
ID      AAZ23428 standard; DNA: 4481 BP.
XX
XX      AAZ23428;
AC      19-JAN-2000 (first entry)
XX
XX      Human SCA2 DNA.
DE
XX      Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KM      huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KM      atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KM      Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KM      dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KM      neoplastic; malignant; autoimmune; fibrotic; ss.
XX
OS      Homo sapiens.

```


XX	Key	Location/Qualifiers
FT	CDS	163..4101
FT		/*tag= a
FT		/product= "SCA2"
PN		MO9945944-A1.
XX		
PD		16-SEP-1999.
XX		
PF		11-MAR-1999; 99WO-US05250.
XX		
PR		12-MAR-1998; 98US-0041886.
XX		
PA		(BURN-) BURNHAM INST.
PI		Bredesen DE, Rabinzadeh S;
XX		
DR		WPI; 1999-561617/47.
XX		
XX		P-PSDB; AAY33495.
PT		New proapoptotic dependence peptides, used to develop products for
PT		treating, e.g. Alzheimer's disease -
XX		
PS		Disclosure: Page 130-135; 199PP: English.
XX		
CC		This invention describes novel pure proapoptotic dependence peptides
CC		which comprise a sequence of an active dependence domain selected from
CC		dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC		huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC		SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC		of inducing cell death and can be used to develop products to mediate or
CC		inhibit apoptosis. The methods can be used for reducing the severity of
CC		a proapoptotic dependence domain mediated pathological conditions e.g.
CC		Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC		Sphincter pupillator ataxia, dentatorubropallidoluysian atrophy,
CC		Machado-Joseph disease, stroke or head trauma. They can also be used for
CC		reducing the severity of a pathological condition mediated by upregulated
CC		cell proliferation or cell survival e.g. neoplastic, malignant,
CC		autoimmune or fibrotic conditions. This sequence encodes the human
CC		SCA2 polypeptide described in the method of the invention.
XX		
SO		Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;
XX		
Query Match	96.8%;	Score 30; DB 20; Length 4481;
Best Local Similarity	100.0%;	Pred. No. 0.072;
Matches 30; Conservative	0;	Mismatches 0; Indels 0; Gaps 0.
OY	1	ctcgcggggcctcccccgccttcgtcgc 30
DB	451	ctcgcggggcctcccccgccttcgtcgc 480
XX		
RESULT 8		
XX		
ID	AAL34906	
XX		
AC	AAL34906;	
XX		
DT	08-JAN-2002	(first entry)
XX		
DE		Human musculoskeletal system related polynucleotide seq ID NO 248.
XX		
KW		Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW		antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW		vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW		cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW		neurological disease; infection; human; secreted protein;
XX		
XX		musculoskeletal system; ss.
XX		
OS	Homo sapiens.	
XX		

[illegible]

PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239353
PR	13-OCT-2000	2000US-0239397
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246674
PR	08-NOV-2000	2000US-0246675
PR	08-NOV-2000	2000US-0246676
PR	08-NOV-2000	2000US-0246677
PR	08-NOV-2000	2000US-0246678
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249246
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250130
PR	05-DEC-2000	2000US-0251031
PR	05-DEC-2000	2000US-0251188
PR	06-DEC-2000	2000US-0251671
PR	06-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251889
PR	08-DEC-2000	2000US-0251930
PR	11-DEC-2000	2000US-0254097
PR	05-JAN-2001	2001US-0259678
XX	(HUMA-)	HUMAN GENOME SCI INC.
PA		
PI	Rosen CA, Barash SC, Ruben SM,	

XX	WP1: 2001-451937/48.
DR	P-PSDB: ABB03324.
XX	Isolated polypeptide for treating, preventing and/or prognosing
PT	disorders related to the musculoskeletal system including
PT	musculoskeletal cancers and also for testing and detection e.g.
PT	diagnosis -
XX	
PS	Claim 1; SEQ ID NO 248; 781pp + Sequence Listing; English.
XX	
CC	The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC	(ABB03087-ABB04109) associated with the musculoskeletal system useful
CC	for preventing, treating or ameliorating medical conditions e.g. by
CC	protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (ant)agonists are useful in the diagnosis, treatment
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC	other cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC	(c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 328 BP; 57 A; 93 C; 113 G; 65 T; 0 other;
XX	
Query Match	69.0%; Score 21.4; DB 22; Length 328;
Best Local Similarity	80.6%; Pred No.78;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
Oy	1 ctcgcgcgcgcctcccccgcctcgcgc 31
Dd	88 ctcgcgcgcctcccccgcctcgcgc 118
XX	
RESULT 9	
AAV62163/c	
ID	AAV62163 standard; DNA; 427 BP.
XX	
AC	AAV62163;
XX	
DT	23-DEC-1998 (first entry)
XX	
DE	HSV-2 strain SB5 Contlig ID 2 DNA sequence.
XX	
KM	HSV-2 strain SB5; immunological response induction; therapy;
KM	antiviral identification; viral protein inhibitor; ss.
XX	
OS	Herpes simplex virus type 2.
XX	
PN	W09820016-A1.
XX	
PD	14-MAY-1998.
XX	
PF	31-OCT-1997; 97WO-US20016.
XX	
PR	09-JUN-1997; 97US-0049018.
XX	
PR	04-NOV-1996; 96US-0030279.
XX	
PA	(SMIK) SMITHKLINE BEECHAM CORP.
XX	
PI	Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
PI	Esset KM, Leary JT;
XX	
DR	WP1: 1998-286847/25.

XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
XX
PS Claim 1: Page 465; 748bp; English.
XX
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC sequence of the invention. This sequence was isolated from HSV-2 strain
CC SB5 (deposited as ATCC VR-2546), is designated Config ID 2. Proteins
CC encoded by the HSV DNA sequences can be used for the treatment or
CC prevention of disease, to induce an immunological response in a mammal or
CC to identify inhibitors, activators or novel antivirals. Antagonists of
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC or a vector containing it can also be used to induce an immunological
CC response in a mammal.
XX
SQ Sequence 427 BP; 46 A; 203 C; 142 G; 36 T; 0 other;

Query Match 66.5%; Score 20.6; DB 19; Length 427;
Best Local Similarity 85.2%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 gcgggcctcccgccctctgcctcc 31
||||| ||||| ||||| ||||| |||||
DB 360 GCGGCGACCCCGCCGCTGCTGCTCC 334

RESULT 10
AAK8967/c
ID AAK8967 standard; DNA; 726 BP.
XX
AC AAK8967;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 3543.
XX
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
OS
PN WO200155314-A2.
PD
XX
XX 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-019874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234474.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236127.
PR 29-SEP-2000; 2000US-0236127.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.

0y	1	ctcggcggcctcccccgccttcgtc	27
db	500	ctcttcggcctccctccgccttcgtc	474
<p>Query Match 66.5%; Score 20.6; DB 22; Length 726; Best Local Similarity 85.2%; Pred.No.1.4e+02; Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;</p>			

```

RESULT 11
AAC56198/c
ID AAC56198 standard; DNA: 1008 BP.
XX
AC AAC56198;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #329.
XX
KW plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
XX
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX
XX WPI: 2000-579369/54.
XX
XX
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide
XX
XX
XX Claim 1; Page 131; 747pp; English.
XX
XX
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX sequence for one such transcription factor. The transcription factor may
XX be used to produce a plant having modified gene expression such as a
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX mahogany species or to modify the activity of a polypeptide in a plant.
XX The transcription factors of the present invention are members from the
XX following families of regulatory proteins: bZIP, bZIP family of G-box
XX binding factors, basic helix-loop-helix zipper,
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
XX and ERBSs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX and MYB.
XX
XX
XX Sequence 1008 BP; 175 A; 315 C; 331 G; 187 T; 0 other:
XX
XX
XX Query Match 66.5%; Score 20.6; DB 21; Length 1008;
XX Best Local Similarity 85.2%; Pred. No. 1.3e+02;
XX Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 ctcgcggcgccctcccgccctgctc 27
XX | ||||| ||| ||||| |||
XX Db 340 CGCGCGGGCCGCCGCCGCCCTCTTC 314
XX
XX
XX RESULT 12
XX AAA09686/c
XX ID AAA09686 standard; DNA: 3957 BP.
XX
XX AC AAA09686;
XX
XX

```

DT 31-JAN-2001 (first entry)
 XX HSV-2 Immediate early protein ICP4 DNA sequence.
 DE
 XX
 XX Herpes-simplex-virus type 2; HSV-2; infected cell protein 4; ICP4;
 KM vaccine; infection; ds.
 XX
 OS Herpes simplex virus type 2.
 XX
 XX W09516779-A1.
 PN
 XX
 PD 22-JUN-1995.
 XX
 PF 13-DEC-1994: 94MO-EP04138.
 XX
 PR 14-DEC-1993: 93GB-0025496.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Pala P, Gheysen DR, Slaoui MM, Koutsoukos MC;
 DR
 DR WPI: 2001-024142/03.
 XX
 DR P-PSDB: AAB26874.
 XX
 PT Immediate early herpes-simplex-virus type 2 (HSV-2) ICP4 protein is
 PT used in vaccines for therapeutically or prophylactically treating HSV
 PT infections -
 XX
 PS Claim 5; Page 16; 28pp; English.
 XX
 CC This invention relates to an immediate early herpes-simplex-virus type 2
 CC (HSV-2) infected cell protein 4 (ICP4) recognised by human cytotoxic T
 CC cells. HSV-2 ICP4 protein is recognized by cytotoxic T-lymphocyte (CTL)
 CC cells in humans and is used in vaccines for therapeutically or
 CC prophylactically treating HSV infections. Pharmaceutical compositions of
 CC HSV-2 ICP4 protein may be used to treat patients suffering from HSV
 CC infections, to prevent or decrease recurrent herpes disease, frequency,
 CC severity and duration of episodes. The present sequence represents HSV-2
 CC DNA encoding ICP4.
 XX
 SQ Sequence 3957 BP; 368 A; 1656 C; 1568 G; 365 T; 0 other:

 Query Match 66.5%; Score 20.6; DB 22; Length 3957;
 Best Local Similarity 85.2%; Pred. No. 1.2e+02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

 OY 5 gcgggacctccgcgcctcgtcgtcc 31
 ||||| ||||| ||||| ||||| |||||
 DB 878 GCGGCGACCCCGCCCTCCTCGTCGCC 852

 RESULT 13
 AAV62175/c
 ID AAV62175 standard; DNA: 16812 BP.
 XX
 AC AAV62175;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 12 DNA sequence.
 XX
 KM HSV-2 strain SB5; Immunological response induction; therapy;
 KM antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 OS
 XX Key Location/Qualifiers
 FT CDS 127..1371
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72159"
 FT CDS complement (1553..2428)

FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72160"
 FT CDS 2714..4159
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72161"
 FT CDS 6835..6948
 FT /tag= d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72162"
 FT CDS 7392..8573
 FT /tag= e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72163"
 FT CDS 8173..9893
 FT /tag= f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in AAW72164"
 FT CDS 10212..11858
 FT /tag= g
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72165"
 FT CDS 12010..12147
 FT /tag= h
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72166"
 FT CDS 12247..12516
 FT /tag= i
 FT /product= "ORF#9 protein"
 FT /note= "encoded protein shown in AAW72167"
 FT CDS complement (13004..13912)
 FT /tag= j
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72168"
 FT CDS 15899..16582
 FT /tag= k
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72169"
 XX
 PN W09820016-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 31-OCT-1997: 97MO-US20016.
 XX
 PR 09-JUN-1997: 97US-0049018.
 XX
 PR 04-NOV-1996: 96US-0030279.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
 PI Esser KM, Leary JT;
 XX
 DR WPI: 1998-286847/25.
 DR P-PSDB: AAW72159, AAW72160, AAW72161, AAW72162, AAW72163, AAW72164,
 DR AAW72165, AAW72166, AAW72167, AAW72168, AAW72169.
 XX
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal
 XX
 PS Claim 1; Page 505-512; 748pp; English.
 XX
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
 CC sequence of the invention. This sequence was isolated from HSV-2 strain
 CC SB5 (deposited as ATCC VR-2546), is designated Contig ID 12, and encodes
 CC 11 HSV-2 proteins. The proteins can be used for the treatment or
 CC prevention of disease, to induce an immunological response in a mammal or
 CC to identify inhibitors, activators or novel antivirals. Antagonists of
 CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
 CC or a vector containing it can also be used to induce an immunological
 CC response in a mammal.

```
XX Sequence 16812 BP; 2708 A; 5989 C; 5367 G; 2748 T; 0 other;
SQ
Query Match 66.5%; Score 20.6; DB 19; Length 16812;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 5 gcgggcctcccgccctcgtcgtcc 31
||||| ||||||| |||||||
Db 16768 GCGGGCACCCCGCCCTCTGCTGCTCC 16742

RESULT 14
AAD25519
ID AAD25519 standard; DNA; 154746 BP.
XX
AC AAD25519;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human herpesvirus 2 complete DNA genome.
XX
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoacide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.
XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension -
XX
PS Disclosure; Page 90-132; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is human herpesvirus 2 complete DNA genome related
CC to the invention.
XX
SQ Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;
```

```
OY 5 gcgggcctcccgccctcgtcgtcc 31
||||| ||||||| |||||||
Db 131155 gcgggcaccccgccctcgtcgtcc 131181

RESULT 15
AAD25519/c
ID AAD25519 standard; DNA; 154746 BP.
XX
AC AAD25519;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human herpesvirus 2 complete DNA genome.
XX
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoacide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US11372.
XX
PR 07-APR-2000; 2000US-195680P.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
DR WPI; 2002-066308/09.
XX
PT Composition for oral delivery of vaccines, comprises expression vector
PT containing antigenic genomic sequence, bound to aggregated
PT protein-polycationic polymer conjugate or suspension -
XX
PS Disclosure; Page 90-132; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
CC bound to an aggregated protein-polycationic polymer conjugate or
CC suspension. The expression vector contains a promoter polynucleotide
CC sequence operatively linked to a polynucleotide sequence encoding an
CC antigen which is a fragment of a gene or genome associated with an
CC infectious disease, cancer and autoimmune disease such as rheumatoid
CC arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
CC consisting of bacterium, fungus, protozoa and virus such as human
CC immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
CC virus (HCV), influenza and respiratory syncytial virus (RSV), and
CC optionally comprising a nucleotide sequence encoding a cytokine (or a
CC cytokine expression vector), is useful for inducing an immune response
CC (systemic and/or mucosal) in an organism. The cytokine expression vector
CC contains a sequence for granulocyte macrophage-colony stimulating factor
CC (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
CC the antigen and the cytokine are under transcriptional control of same or
CC different promoter polynucleotide sequences. The expression vector, as a
CC DNA vaccine is useful for treating a condition in an organism. The
CC present sequence is human herpesvirus 2 complete DNA genome related
CC to the invention.
XX
SQ Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;
```

Query Match 66.5%; Score 20.6; DB 24; Length 154746;
Best Local Similarity 85.2%; Pred. No. 89;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 gcgggctcccccgccttcgtctcc 31
||||| ||||| |||||
DB 150587 gcgggcaccccccctcctcgtctcc 150561

Search completed: August 14, 2002, 22:06:32
Job time: 11687 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:51:00 ; Search time 203.42 Seconds

(without alignments)
37,433 Million cell updates/sec

Title: US-09-707-919-3

Perfect score: 31

Sequence: 1 ctgcgcggcgtcccccgttcgtcgc 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_patents.NA:*
2: /cgn2_6/plodata/1/lna/5A.COMB.seq:*
3: /cgn2_6/plodata/1/lna/5B.COMB.seq:*
4: /cgn2_6/plodata/1/lna/6A.COMB.seq:*
5: /cgn2_6/plodata/1/lna/6B.COMB.seq:*
6: /cgn2_6/plodata/1/lna/PCTUS.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	96.8	355	4	US-09-043-303-1
2	30	96.8	623	4	US-09-043-303-5
3	30	96.8	4481	4	US-09-041-886-18
4	19.4	62.6	1200	1	US-08-356-397-1
5	18.8	60.6	1741	1	US-08-565-655-5
6	18.6	60.0	1962	4	US-08-791-115B-3
7	18.4	59.4	2936	2	US-08-714-677-10
8	18.4	59.4	2936	2	US-08-393-540-10
9	18.4	59.4	2936	2	US-08-714-537-10
10	18.4	59.4	3196	2	US-09-096-982-4
11	18.4	59.4	3196	2	US-08-653-650A-4
12	18.2	58.7	696	4	US-08-998-416-1050
13	18.2	58.7	1650	2	US-08-743-637B-172
14	18.2	58.7	1650	3	US-08-526-840B-172
15	18.2	58.7	3051	1	US-08-241-766-10
16	18	58.1	1743	3	US-09-032-365A-18
17	18	58.1	1785	3	US-08-729-416C-8
18	18	58.1	2316	6	5258283-6
19	18	58.1	2575	4	US-09-077-354B-1
20	18	58.1	10380	4	US-09-077-354B-3
21	18	58.1	4411529	4	US-09-103-840A-1
22	17.8	57.4	678	4	US-09-459-956-6
23	17.8	57.4	3000	1	US-08-393-985-3
24	17.8	57.4	3323	1	US-07-980-528-1
25	17.8	57.4	28604	2	US-08-592-874-1
26	17.8	57.4	28804	3	US-09-096-942-2
27	17.8	57.4	28804	3	US-09-096-867-2

ALIGNMENTS

c 28	17.8	57.4	43804	4	US-09-171-461-1	Sequence 1, Appl1
c 29	17.8	57.4	71989	4	US-09-443-501A-2	Sequence 2, Appl1
c 30	17.8	57.4	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
c 31	17.8	57.4	4411529	4	US-09-103-840A-1	Sequence 6, Appl1
c 32	17.6	56.8	2588	2	US-08-796-414B-6	Sequence 1, Appl1
c 33	17.4	56.1	4405	2	US-07-885-972A-3	Sequence 3, Appl1
c 34	17.4	56.1	4405	2	US-08-745-880-3	Sequence 3, Appl1
c 35	17.4	56.1	4405	2	US-08-480-382-3	Sequence 3, Appl1
c 36	17.4	56.1	4978	1	US-08-220-603A-1	Sequence 1, Appl1
c 37	17.4	56.1	49272	1	US-08-614-770A-1	Sequence 1, Appl1
c 38	17.4	56.1	50341	1	US-08-247-901C-1	Sequence 1, Appl1
c 39	17.4	56.1	50341	2	US-09-075-901C-1	Sequence 1, Appl1
c 40	17.4	56.1	52297	4	US-09-426-436-1	Sequence 1, Appl1
c 41	17.4	56.1	52297	4	US-08-705-557-1	Sequence 1, Appl1
c 42	17.4	56.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
c 43	17.2	55.5	884	2	US-08-901-200A-11	Sequence 11, Appl1
c 44	17.2	55.5	884	3	US-09-219-391-11	Sequence 11, Appl1
c 45	17.2	55.5	1100	2	US-08-776-210-4	Sequence 4, Appl1

```

RESULT 1
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
US-09-043-303-1

```

```

Query Match          96.8%; Score 30; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctgcgcggcgtcccccgttcgtcgc 30
Db 149 ctgcgcggcgtcccccgttcgtcgc 178

RESULT 2
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17

```


SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 5
LENGTH: 623
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (341)..(583)
FEATURE:
OTHER INFORMATION: TSP-2
US-09-043-303-5

Query Match 96.8%; Score 30; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcgagcgagcctcccgcccttcgtcgc 30
|||||
Db 149 ctcgagcgagcctcccgcccttcgtcgc 178

RESULT 3
US-09-041-886-18
Sequence 18, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Shantoz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

Query Match 96.8%; Score 30; DB 4; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ctcgagcgagcctcccgcccttcgtcgc 30

|||||
Db 451 CTCGGCGGCTCTCCCGCCCTTCGTCGTC 480

RESULT 4
US-08-356-397-1/C
Sequence 1, Application US/08356397
Patent No. 5648259
GENERAL INFORMATION:
APPLICANT: Maillet, Jacques
APPLICANT: Smirnova, Tania
TITLE OF INVENTION: NOVEL POLYPEPTIDES HAVING NMDA RECEPTOR
TITLE OF INVENTION: ACTIVITY, NUCLEIC ACIDS ENCODING SAID POLYPEPTIDES AND
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,397
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Julie K.
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: ST92038-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1200 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 211..1077
US-08-356-397-1

Query Match 62.6%; Score 19.4; DB 1; Length 1200;
Best Local Similarity 79.3%; Pred. No. 51;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 cgcgcgagcctcccgcccttcgtcgc 31
|||||
Db 206 CGACGGCTCTCTCCCTTCGTCGTC 178

RESULT 5
US-08-565-655-5/C
Sequence 5, Application US/08565655
Patent No. 5688939
GENERAL INFORMATION:
APPLICANT: Potter, Sharon L
APPLICANT: Ward, Eric R
TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
TITLE OF INVENTION: DNA Coding Therefor
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

```

1 ADDRESS: Ciba Patent Department
2 STREET: 540 White Plains Rd., POB 2005
3 City: Tarrytown
4 STATE: NY
5 COUNTRY: USA
6 ZIP: 10591-9005
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: Patentin Release #1.0, Version #1.30B
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/565,655
13 FILING DATE:
14 CLASSIFICATION: 210
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER:
17 FILING DATE: 12-DEC-1994
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Elmer, James Scott
20 REGISTRATION NUMBER: 36,129
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (919) 541-8614
23 TELEFAX: (919) 541-8689
24 INFORMATION FOR SEQ ID NO: 5:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 1741 base pairs
27 TYPE: nucleic acid
28 STRANDEDNESS: single
29 TOPOLOGY: linear
30 MOLECULE TYPE: cDNA
31 HYPOTHETICAL: NO
32 FEATURE:
33 NAME/KEY: CDS
34 LOCATION: 1..1428
35 OTHER INFORMATION: /product= "Wheat Adenylosuccinate
36 OTHER INFORMATION: Synthetase"
37
38 US-08-565-655-5
39
40 Query Match 60.6%; Score 18.8; DB 1; Length 1741;
41 Best Local Similarity 76.7%; Pred. No. 81;
42 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0.
43
44 QY 2 tcggcgagctctcccgcccttcgtcgcc 31
45 ||| ||| ||| ||| ||| ||| ||| ||| |||
46 Db 215 TCGACGAGCTTCCCTCCCTCCCTCGTCCGCC 186
47
48 RESULT 6
49 US-08-791-115B-3
50 Sequence 3, Application US/08791115B
51 Patent No. 6262242
52 GENERAL INFORMATION:
53 APPLICANT: Steck, Peter
54 APPLICANT: Pershouse, Mark A.
55 APPLICANT: Jasser, Samar
56 APPLICANT: Yung, W.K. Alfred
57 APPLICANT: Tavetigian, Sean V
58 TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
59 NUMBER OF SEQUENCES: 27
60 CORRESPONDENCE ADDRESSES:
61 ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
62 STREET: 555 Thirteenth Street, N.W., Suite 701-E
63 CITY: Washington
64 STATE: DC
65 COUNTRY: USA
66 ZIP: 22204
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS
71 SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/791,115B
FILING DATE: 30-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 38,957
REFERENCE/DOCKET NUMBER: 2318-134.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-683-6040
TELEFAX: 202-683-7031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-791-115B-3

Query Match          60.0%; Score 18.6; DB 4; Length 1962;
Best Local Similarity 84.0%; Pred. NO. 95;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3  cggcgagcctcccgcccttcgic 27
          |||||  |||  ||  |||||
Db      115  CGCGCGCCTCGCCTCGTCGTCGTC 139

RESULT      7
US-08-714-677-10/C
Sequence 10, Application US/08714677
Patent No. 5871977
GENERAL INFORMATION:
APPLICANT: KUBOTA, Michio
APPLICANT: TSUSAKI, Kenji
APPLICANT: MARUTA, Kazuhiko
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,677
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/393,540
FILING DATE: 23-FEB-1995
APPLICATION NUMBER: JP 090728
FILING DATE: 06-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 047956
FILING DATE: 23-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 047940
FILING DATE: 23-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 090705
FILING DATE: 06-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.

```

```

? REGISTRATION NUMBER: 25,618
? REFERENCE/DOCKET NUMBER: KUBOTA-4
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-628-5197
? TELEFAX: 202-737-3528
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2936 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 565..2880
? US-08-714-677-10

```

	Query Match	59.4%	Score 18.4	DB 2	length 2936
	Best Local Similarity	78.6%	Pred. No. 1.e+02		
	Matches 22:	Conservative	0;	Mismatches 6;	Indels ?; Gaps 0;
Oy	1 ctgcggagcctccgccccccttcgtg	28			
Db	505 cccccggcctttgcggccgcctctcgtc	478			

US-08-393-540-10/c
Sequence 10, Application US/08393540
Patent No. 5871993
GENERAL INFORMATION:
APPLICANT: KUBOTA, Michio
APPLICANT: TSUSAKI, Kenji
APPLICANT: MARUTA, Kazuhiko
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,540
FILING DATE: 23-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 090728
FILING DATE: 06-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 047956
FILING DATE: 23-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 047940
FILING DATE: 23-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 090705
FILING DATE: 06-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: KUBOTA-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197

```

? TELEFAX: 202-737-3528
? INFORMATION FOR SEQ ID NO: 10
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2936 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 565..2880
?
US-08-393-540-10

```

Query Match	59.48%	Score 18.4	DB 2	Length 2936
Best Local Similarity	78.68%	Pred. No. 1.1e+02		
Matches 22	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Oy	1	ctcgagcgagctccgcgcctctatcg	28	
Db	505	CCCCGCGAGCTTCCGCCCGCCTGCTG	478	

```

1      RESULT          9
2      US-08-714-537-10/C
3      : Sequence 10, Application US/08714537
4      : Patent No. 5871994
5      :
6      : GENERAL INFORMATION:
7      :   APPLICANT: KUBOTA, Michio
8      :   APPLICANT: TSUSAKI, Kenji
9      :   APPLICANT: MARUTA, Kazuhiko
10     :   APPLICANT: SUGIMOTO, Toshiyuki
11     :   TITLE OF INVENTION: DNA ENCODING ENZYME, RECOMBINANT DNA AND
12     :   TITLE OF INVENTION: ENZYME, TRANSFORMANT, AND THEIR PREPARATIONS AND USES
13     :   NUMBER OF SEQUENCES: 17
14     :   CORRESPONDENCE ADDRESS:
15     :     ADDRESSEE: BROWDY AND NEWMARK, P. L. L. C.
16     :     STREET: 419 Seventh Street, N.W., Suite 400
17     :     CITY: Washington
18     :     STATE: D.C.
19     :     COUNTRY: USA
20     :     ZIP: 20004
21     :
22     : COMPUTER READABLE FORM:
23     :   MEDIUM TYPE: Floppy disk
24     :   COMPUTER: IBM PC compatible
25     :   OPERATING SYSTEM: PC-DOS/MS-DOS
26     :   SOFTWARE: PatentIn Release #1.0, Version #1.30
27     :   CURRENT APPLICATION DATA:
28     :     APPLICATION NUMBER: US/08/714,537
29     :     FILING DATE:
30     :       CLASSIFICATION:
31     :     PRIOR APPLICATION DATA:
32     :       APPLICATION NUMBER: US/08/393,540
33     :       FILING DATE: 23-FEB-1995
34     :       APPLICATION NUMBER: JP 090728
35     :       FILING DATE: 06-APR-1994
36     :       PRIOR APPLICATION DATA:
37     :         APPLICATION NUMBER: JP 047956
38     :         FILING DATE: 23-FEB-1994
39     :         PRIOR APPLICATION DATA:
40     :           APPLICATION NUMBER: JP 047940
41     :           FILING DATE: 23-FEB-1994
42     :           PRIOR APPLICATION DATA:
43     :             APPLICATION NUMBER: JP 090705
44     :             FILING DATE: 06-APR-1994
45     :             ATTORNEY/AGENT INFORMATION:
46     :               NAME: BROWDY, Roger L.
47     :               REGISTRATION NUMBER: 25,618
48     :               REFERENCE/DOCKET NUMBER: KUBOTA-4
49     :               TELECOMMUNICATION INFORMATION:
50     :                 TELEPHONE: 202-628-5197
51     :                 TELEFAX: 202-737-3528
52     :   INFORMATION FOR SEQ ID NO: 10:

```


APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264rtis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NO. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCCI976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO: 1050:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAGI640UP
US-08-998-416-1050

Query Match 58.7%; Score 18.2; DB 4; Length 696;
Best local Similarity 74.2%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 ctcgagcgagcctcccgcccttcgtctcc 31
||||| ||| ||||| ||| ||||| |||
Db 391 CTCGTCGTGGAGACCCGCGCGTCGTCTAC 361

RESULT 13
US-08-743-637B-172/c
Sequence 172, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA

ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414)277-5591
INFORMATION FOR SEQ. ID NO: 172:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-743-637B-172

Query Match 58.7%; Score 18.2; DB 2; Length 1650;
Best local Similarity 74.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 ctcgagcgagcctcccgcccttcgtctcc 31
||||| ||| ||||| ||| ||||| |||
Db 1519 CTYGGCGGACATCCACGCCGCGTCGTCTC 1489

RESULT 14
US-08-526-840B-172/c
Sequence 172, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.

```

?      REGISTRATION NUMBER: 35/433
?      REFERENCE/DOCKET NUMBER: 850586.90012
?
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (414) 277-5000
?      TELEFAX: (414) 277-5591
?      INFORMATION FOR SEQ ID NO: 172:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1650 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: double
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
US-08-526-8408-172

```

Query Match	58.7%	Score 18.2;	DB 3;	Length 1650;
Best Local Similarity	74.2%	Pred. No. 1.3e+02;		
Matches 23; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0

```

QY      1  ctgcgcgcgcctccccgccttcgtctcc 31
          || ||||| | ||| ||||| | ||||| |
Db 1519  CTTGGCGGACATCCACGCCGACGCTGTGC 1489

```

RESULT 15
US-08-241-766-10/c
: Sequence 10. Application US/08241766
: Patent No. 5686590
: GENERAL INFORMATION:
: APPLICANT: JACOBS, W. R.
: APPLICANT: COLLINS, D. M.
: APPLICANT: BANERJEE, A.
: APPLICANT: DELISTE, G. W.
: APPLICANT: WILSON, T. M.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
: TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 Page Mill Road
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/241,766
: FILING DATE: 12-May-1994
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: MONROY, GLADYS H.
: REGISTRATION NUMBER: 32,430
: REFERENCE/DOCKET NUMBER: 25237-20003.20
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 813-5600
: TELEFAX: (415) 494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3051 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IS-08-241-766-10

Query Match	58.7%	Score 18.2;	DB 1;	Length 3051;
Best Local Similarity	74.2%	Pred. No. 1.3e+02;		
Matches 23; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

QY	1	ctcgagcgggacctccccgcgcccttcgtcc	31
Db	2562	CTCGGCGGCCTGCTGGCAGTTGTTGTCC	2532

```
Search completed: August 14, 2002, 21:51:15
Job time: 13508 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:06 ; Search time 7749.14 Seconds
(without alignments)
53.994 Million cell updates/sec

Title: US-09-707-919-3

Perfect score: 31
Sequence: 1 ctcgcgcgcctcccccgccttcgcctcc 31

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estluc:*
5: em_estluc:*
6: em_estluc:*
7: em_estluc:*
8: em_estluc:*
9: em_estluc:*
10: em_estluc:*
11: em_estluc:*
12: em_estluc:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	1100	10	BM455214
2	30	96.8	482	9	AL039573
3	30	96.8	500	10	BI547486
4	26.4	85.2	364	10	BE457923
5	22	71.0	900	10	BC309511
6	22	71.0	1256	12	AG032225
7	21.6	66.7	343	10	BI478400
8	21.6	66.7	897	12	CNS032CV
9	21.6	66.7	1030	12	AG747830
10	21.4	69.0	50	9	AU102974
11	21.4	69.0	736	12	AG068851
12	21.4	69.0	1082	12	AG064664
13	21.4	69.0	1585	10	BF206104
14	21	67.7	570	9	AV588796
15	21	67.7	701	12	AG076621
16	21	67.7	884	10	AL535465
17	21	67.7	906	10	BM459522

c	18	21	67.7	1602	10	BG843193
c	19	20.6	66.5	252	10	BM443926
c	20	20.6	66.5	293	10	BI780637
c	21	20.6	66.5	313	9	AV933796
c	22	20.6	66.5	334	10	BM374143
c	23	20.6	66.5	437	10	BM100923
c	24	20.6	66.5	441	9	AI253086
c	25	20.6	66.5	496	10	BG417150
c	26	20.6	66.5	570	9	AV933056
c	27	20.6	66.5	582	9	AV932088
c	28	20.6	66.5	591	9	AV934263
c	29	20.6	66.5	731	10	BC344472
c	30	20.6	66.5	936	9	AL571687
c	31	20.4	65.8	283	10	BE575070
c	32	20.4	65.8	287	9	BB500108
c	33	20.4	65.8	319	12	FR0032847
c	34	20.4	65.8	418	10	BI165311
c	35	20.4	65.8	430	10	BI166598
c	36	20.4	65.8	445	10	BI169120
c	37	20.4	65.8	446	10	BE586264
c	38	20.4	65.8	474	12	TA370F10P
c	39	20.4	65.8	485	10	BI304368
c	40	20.4	65.8	490	9	AI533013
c	41	20.4	65.8	501	12	AQ850584
c	42	20.4	65.8	512	10	BI177625
c	43	20.4	65.8	527	10	BI237644
c	44	20.4	65.8	529	10	BG817511
c	45	20.4	65.8	533	10	BI162612

ALIGNMENTS

RESULT 1
BM455214
LOCUS
DEFINITION
AGENCOURT 6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
5' mRNA sequence.

ACCESSION
BM455214
VERSION
BM455214.1 GI:18504254
KEYWORDS
EST.

SOURCE
ORGANISM
human.

REFERENCE
1 (bases 1 to 1100)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILU)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILU at:
http://image.llnl.gov
Plate: LILU12134 row: k column: 12
High quality sequence stop: 623.
Location/Qualifiers

FEATURES
source
1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/issue="NIH-MGC_85"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph. Vector: pCMV-SPORE6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

```

BASE COUNT      240 a      329 c      306 g      219 t      6 others
ORIGIN
Query Match      100.0%; Score 31; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ctcggcgagcctccccccttcgtcgc 31
|||||
Db 72 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 102

RESULT 2
AL039573      482 bp      mRNA      linear      EST 29-FEB-2000
LOCUS      DKFZP434D311.1 434 (synonym: hces3) Homo sapiens cDNA clone
DEFINITION      DKFZP434D311.5, mRNA sequence.
ACCESSION      AL039573
VERSION      AL039573.1 GI:5408612
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 482)
AUTHORS      Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann
, S.
TITLE      EST (Duesterhoeft, et al.)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Duesterhoeft A
MIPS      Am Klopfersplitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZP434D311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heudnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
FEATURES
source
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="434 (synonym: hces3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT      49 a      218 c      145 g      70 t

ORIGIN
Query Match      96.8%; Score 30; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ctcggcgagcctccccccttcgtcgc 30
|||||
Db 98 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 127

RESULT 3
BI547486      500 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      B63191091F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:526235 5',
DEFINITION      mRNA sequence.
ACCESSION      BI547486
VERSION      BI547486.1 GI:15434798
KEYWORDS      EST.

```

```

SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 500)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NIGRI), Shiraiki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM11661 row: e column: 24
High quality sequence stop: 485.
Location/Qualifiers
FEATURES
source
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:526235"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (ctcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NIGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT      57 a      222 c      150 g      71 t

ORIGIN
Query Match      96.8%; Score 30; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ctcggcgagcctccccccttcgtcgc 30
|||||
Db 101 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 130

RESULT 4
BE457923      364 bp      mRNA      linear      EST 26-JUL-2000
LOCUS      US99C12.x1 Soares-thymus_2bmt Mus musculus cDNA clone
DEFINITION      IMAGE:3326518 3', similar to TR:070305 070305 SPINOCEREBELLAR ATAXIA
2 HOMOLOG ;, mRNA sequence.
ACCESSION      BE457923
VERSION      BE457923.1 GI:9480561
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 364)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1070682

```


LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..1256
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-006G11.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 363 a 359 c 480 g 17 t 37 others
ORIGIN

Query Match 71.0%; Score 22; DB 12; Length 1256;
Best Local Similarity 80.6%; Pred. No. 3e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 1 ctcggcgagctccccccttcgtcgtc 31
Db 448 CTCGGGGGCCCTCTCTCTCTCTC 418

RESULT 7
BI478400 343 bp mRNA linear EST 27-AUG-2001
LOCUS 949065D08.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION BI478400
VERSION BI478400.1 GI:15312818
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 343)
Walbot V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949065 row: D Column: 08.
Location/Qualifiers
1..343
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridz vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf

primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."
BASE COUNT 34 a 131 c 138 g 40 t
ORIGIN

Query Match 69.7%; Score 21.6; DB 10; Length 343;
Best Local Similarity 85.7%; Pred. No. 4e+03;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 3 cggcgagctccccccttcgtcgtc 30
Db 126 CGGCGGGGCCCTCTCTCTCTCTC 153

RESULT 8
CNS032CV
LOCUS 897 bp DNA linear GSS 18-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone
070A02 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL267448 GI:7989256
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 897)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Unpublished
2 (bases 1 to 897)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brotier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 897)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.
Location/Qualifiers
1..897
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="070A02"
/clone_lib="G"
/note="Genoscope sequence ID : C08G070BA01SP1-end :
PUC-ori"
BASE COUNT 214 a 223 c 263 g 183 t 14 others
ORIGIN

Query Match 69.7%; Score 21.6; DB 12; Length 897;
Best Local Similarity 80.0%; Pred. No. 3.9e+03;
Matches 24; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 2 tcggcgagctccccccttcgtcgtc 31
Db 776 TCRGCGGCCCTCTCTCTCTCTC 747

```

RESULT      9
LOCUS       A0747830/c
DEFINITION  A0747830 1030 bp DNA linear GSS 19-JUL-1999
              HS 5537 AL F03 SP6 RPCI-11 Human Male BAC Library Homo sapiens
              genomic clone Plate-1113 Col-5 Row-K, DNA sequence.
ACCESSION   A0747830
VERSION     A0747830.1 GI:5534988
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 1030)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,D., Young,D., Zhao,S., Adams,M.D. and
              Hood,L.
              Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
              Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
JOURNAL     Contact: Mahairas GG, Wallace JC, Hood L
MEDLINE     High Throughput Sequencing Center
COMMENT      University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieterdejong.med.buffalo.edu). Clones may be purchased from
              BACRAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
              or from Research Genetics (info@resgen.com). BAC end Web Server:
              http://www.htsc.washington.edu
              Plate: 1113 row: K column: 5
              Seq primer: SP6
              Class: BAC ends
              High quality sequence stop: 1030.
FEATURES
  source
    1..1030
    Location/Qualifiers
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate-1113 Col-5 Row-K"
      /clone_1lb="RPCI-11 Human Male BAC Library"
      /sex="male"
      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
      Male blood DNA was isolated from one randomly chosen donor
      and partially digested with a combination of EcoRI and
      EcoRI Methylase. Size selected DNA was cloned into the
      pBACe3.6 vector at EcoRI sites"
BASE COUNT  268 a 296 c 402 g 50 t 14 others
ORIGIN
Query Match 69.7%; Score 21.6; DB 12: Length 1030;
Best Local Similarity 82.8%; Pred.No.3.9e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2 tcggcgggccctcccgccctcgctgc 30
Db 434 TCGGGCGNGCTCCCGCCCTTCGGGCC 406

```

```

RESULT      11
LOCUS       A0688851
DEFINITION  A0688851 736 bp DNA linear GSS 01-JUL-1999
              nbx0078002f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION   A0688851
VERSION     A0688851.1 GI:5330019
KEYWORDS    GSS.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 736)
AUTHORS     Wing,R.A. and Dean,R.A.
              A BAC End Sequencing Framework to Sequence the Rice Genome
              Unpublished (1998)
JOURNAL     Contact: Wing RA
COMMENT      Clemson University Genomics Institute
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: rwing@clemson.edu
              Seq primer: TAATACGACTCACTATAGG
              Class: BAC ends
              High quality sequence stop: 81.
FEATURES
  source
    1..736
    Location/Qualifiers
      /organism="Oryza sativa"
      /strain="Japonica"
      /cultivar="Nipponbare"
      /db_xref="taxon:4530"
      /clone="nbx0078002f"
      /clone_1lb="CUGI Rice BAC Library"
      /tissue-type="leaf"
      /lab_host="E. coli DH10B"

```

/note='Vector: pBel0AC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially in those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arunaganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening.'

Query Match	69.0%	Score 21.4	DB 12	length 736
Best Local Similarity	80.6%	Pred. No. 4.5e+03		
Matches	25	Conservative	0	Mismatches 6; Indels 0; Gaps 0
QY	1	ctcggcgagggccctcccccgccttcgtatgctc	31	
DB	520	ccccggcgtcccttcgcgcgtcccttcgcgtcc	550	

RESULT	12
AG064664/c	
LOCUS	AG064664
DEFINITION	1082 bp DNA linear GSS 03-NOV-2001
ACCESSION	Pan troglodytes DNA, clone: PTB-053M24.R, genomic survey sequence.
VERSION	AG064664
KEYWORDS	AG064664.1 GI:16616466
SOURCE	GSS: GSS (genome survey sequence). Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
ORGANISM	BAC library clone: PTB-053M24.R. Pan troglodytes

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1 (sites)	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.	BAC end sequences of library PTB	Unpublished	2 (bases 1 to 1082)	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
			Direct Submission		
			Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:shimobegsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/, Tel:+81-45-503-9111, Fax:81-45-503-9170)		
			Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		

```
FEATURES
  source      location/qualifiers
1. 1092
   /organism="Pan troglodytes"
   /db_xref="taxon:9598"
   /clone="PTB-053M24.R"
   /sex="male"
```

BASE COUNT	ORIGIN	/cell_type="lymphoblast"	/clone_id="PB8 Chimpanzee Male BAC Library"
288 a	222 c	423 g	137 t 12 others

Query Match	69.0%;	Score 21.4;	DB 12;	Length 1082;
Best Local Similarity	60.6%;	Pred. No. 4.5e+03;		
Matches	25;	Conservative	0;	Mismatches 6;
			Indels	0;
Gaps				0;

RESULT 13					
BF206104/c					
LOCUS	BF206104	1585 bp	mrna	linear	EST 06-NOV-2000
DEFINITION	601869458F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4098015 5',				
	mrna sequence.				
ACCESSION	BF206104				

VERSION	BF206104.1	GI:11099690
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 1585)	NIH-MGC	http://mgc.ncl.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.			

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC9364 row: d column: 16
High quality sequence stop: 592.

FEATURES	Location/Qualifiers
source	1. .1585

```

) /ab, host="PH10B (phage-resistant)".
) /note="Organ: brain; Vector: pORB7. Site: 1; XhoI: Site: 2
) EcoRI. cDNA made by oligo-dT priming. Directionally
) cloned into EcoRI/XhoI sites using the following 5'
) adaptor: GCCACGAC(G). Library constructed by Ling Hong
) in the laboratory of Gerald M. Rubin (University of
) California, Berkeley) using ZAP-cDNA synthesis kit
) (Stratagene) and Superscript II RT (Life Technologies).
) Note: this is a NIH_MGC Library."
BASE COUNT      434 a      396 c      557 g      198 t

```

Query Match	69.0%	Score 21.4	DB 10	Length 1585
Best local Similarity	80.6%	Pred. No. 4.4e+03		
Matches 25	Conservative 0	Mismatches 6	Indels 0	Gaps 0
OY	1	ctcgcgcgcctcccccgccttcgtcgc 31		
Db	1387	ctcgcgcgtccttcgcgcgccctccgcctcgc 1357		
RESULT 14				
AV588796/c	AV588796	570 bp	mrna	linear
LOCUS				EST 27-NOV-2001

```

DEFINITION AV588796 Bos taurus brain fetus Bos taurus cDNA clone E1BR02E08
ACCESSION AV588796
VERSION AV588796.1 GI:9699789
KEYWORDS EST.
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 570)
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Mishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..570
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1BR02E08"
/clone_lib="Bos taurus brain fetus"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/notes="Vector: pZ1; Site_1: SacI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 101 a 171 c 175 g 123 t
ORIGIN

Query Match 67.7%; Score 21; DB 9; Length 570;
Best Local Similarity 82.8%; Pred. No. 5.9e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cggcgggcctcccccctcgctgc 31
|||||
Db 248 CGCGGGCCTCCGCCCTCTCTTC 220

RESULT 15
AG076621 701 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-070003.F, genomic survey sequence.
ACCESSION AG076621
VERSION AG076621.1 GI:16628423
GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-070003.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 701)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

```

and Chemical Research (RIKEN), Genomic Sciences Center (GSC):
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB. This BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.

```

PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source
1..701
Location/Qualifiers
1..701
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-070003.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 34 a 300 c 205 g 140 t
ORIGIN

Query Match 67.7%; Score 21; DB 12; Length 701;
Best Local Similarity 82.8%; Pred. No. 5.9e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 cggcgggcctcccccctcgctgc 31
|||||
Db 337 CGCGGGCCTCTCCGCCCTCTCTTC 365

Search completed: August 14, 2002, 21:04:14
Job time: 11002 sec

```


OY	1	ctcgagggcctccgcccttcgtgcg	31
Db	149	CTCGGCGGGCCTCCGCCCTTCGTGC	179
RESULT	2		
LOCUS	ARI59558	572 bp	DNA
DEFINITION	Sequence 18 from patent US 6251589.		linear PAT 17-OCT-2001
ACCESSION	ARI59558		
VERSION	ARI59558.1 GI:16222251		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 572)		
AUTHORS	Tsuji,S. and Sanpei,K.		
TITLE	Method for diagnosing spinocerebellar ataxia type 2 and primers therefor		
JOURNAL FEATURES	Patent: US 6251589-A 18 26-JUN-2001;		
source	Location/Qualifiers 1..572 /organism="unknown"		
BASE COUNT	34 a 277 c 174 g	85 t	2 others
ORIGIN			
Query Match	100.0%; Score 31; DB 6; Length 572;		
Best Local Similarity	100.0%; Pred. No. 2.9;		
Matches	31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1	ctcgagggcctccgcccttcgtgcg	31
Db	149	CTCGGCGGGCCTCCGCCCTTCGTGC	179
RESULT	3		
LOCUS	ARI59546	623 bp	DNA
DEFINITION	Sequence 5 from patent US 6251589.		linear PAT 17-OCT-2001
ACCESSION	ARI59546		
VERSION	ARI59546.1 GI:16222229		
KEYWORDS	.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 623)		
AUTHORS	Tsuji,S. and Sanpei,K.		
TITLE	Method for diagnosing spinocerebellar ataxia type 2 and primers therefor		
JOURNAL FEATURES	Patent: US 6251589-A 5 26-JUN-2001;		
source	Location/Qualifiers 1..623 /organism="unknown"		
BASE COUNT	55 a 292 c 189 g	85 t	2 others
ORIGIN			
Query Match	100.0%; Score 31; DB 6; Length 623;		
Best Local Similarity	100.0%; Pred. No. 2.8;		
Matches	31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1	ctcgagggcctccgcccttcgtgcg	31
Db	149	CTCGGCGGGCCTCCGCCCTTCGTGC	179
RESULT	4		
LOCUS	HSDANSKA2	4163 bp	mRNA
			linear PRI 09-JAN-1997

[illegible]

REFERENCE Mammalia: Eutheria: Primates: Catarrhini: Cercopitheciidae;
AUTHORS 1 (bases 1 to 264)
TITLE Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
JOURNAL Brahmachari, S.K.
PUBMED CAG repeat instability at SCA2 locus: anchoring CAA interruptions
11689490 Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

REFERENCE 2 (bases 1 to 264)
AUTHORS Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..264
/organism="Papio hamadryas"
/db_xref="taxon:9557"
<1..>264
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 25 a 130 c 78 g 31 t

ORIGIN

Query Match 96.8%: Score 30; DB 9; Length 264;
Best Local Similarity 100.0%: Pred. No. 6.9;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctcggcgagcctccgccctcgtcgc 30
|||||
Db 3 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 32

RESULT 9
AF330033 322 bp DNA linear PRI 08-NOV-2001
LOCUS Macaca radiata SCA2 gene, partial sequence.
DEFINITION AF330033
ACCESSION AF330033
VERSION AF330033.1 GI:12382835
KEYWORDS bonnet macaque.
SOURCE Macaca radiata
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
REFERENCE 1 (bases 1 to 322)
AUTHORS Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
TITLE Brahmachari, S.K.
JOURNAL CAG repeat instability at SCA2 locus: anchoring CAA interruptions
PUBMED and linked single nucleotide polymorphisms
11689490 Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

REFERENCE 2 (bases 1 to 322)
AUTHORS Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..322
/organism="Macaca radiata"
/db_xref="taxon:9548"
<1..>322
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 32 a 155 c 95 g 40 t

ORIGIN

Query Match 96.8%: Score 30; DB 9; Length 322;
Best Local Similarity 100.0%: Pred. No. 6.6;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctcggcgagcctccgccctcgtcgc 30
|||||
Db 26 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 55

RESULT 10
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS Presbytis entellus SCA2 gene, partial sequence.
DEFINITION AF330030
ACCESSION AF330030
VERSION AF330030.1 GI:12382832
KEYWORDS Hanuman langur.
SOURCE Presbytis entellus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Colobinae; Presbytis.
REFERENCE 1 (bases 1 to 384)
AUTHORS Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
TITLE Brahmachari, S.K.
JOURNAL CAG repeat instability at SCA2 locus: anchoring CAA interruptions
PUBMED and linked single nucleotide polymorphisms
11689490 Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

REFERENCE 2 (bases 1 to 384)
AUTHORS Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..384
/organism="Presbytis entellus"
/db_xref="taxon:9574"
<1..>384
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 46 a 178 c 109 g 51 t

ORIGIN

Query Match 96.8%: Score 30; DB 9; Length 384;
Best Local Similarity 100.0%: Pred. No. 6.4;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctcggcgagcctccgccctcgtcgc 30
|||||
Db 3 CTCGGCGGCGCTCCCGCCCTTCGTCGTC 32

RESULT 11
AF330029 409 bp DNA linear PRI 08-NOV-2001
LOCUS Gorilla gorilla SCA2 gene, partial sequence.
DEFINITION AF330029
ACCESSION AF330029
VERSION AF330029.1 GI:12382831
KEYWORDS gorilla.
SOURCE Gorilla gorilla
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
REFERENCE 1 (bases 1 to 409)
AUTHORS Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
TITLE Brahmachari, S.K.
JOURNAL CAG repeat instability at SCA2 locus: anchoring CAA interruptions
PUBMED and linked single nucleotide polymorphisms
11689490 Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

REFERENCE 2 (bases 1 to 409)
AUTHORS Choudhry, S. and Brahmachari, S.K.
TITLE Direct Submission

JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES Location/Qualifiers
source 1..409
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
gene <1..>409
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 35 a 196 c 120 g 58 t

ORIGIN

Query Match 96.8%; Score 30; DB 9; Length 409;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcgcgagcctcccccctcctgcgc 30
|||||
Db 31 CTCGGCGGCTCCCGCCCTGTCGTC 60

RESULT 12
AC004085 231758 bp DNA linear HTG 06-NOV-2000
LOCUS Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered
DEFINITION pieces.
AC004085
AC004085.6 GI:11079383
VERSION HTG: HTGS_PHASE1: HTGS_DRAFT.
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 231758)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blinaga,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morjan,M., Morris,S., Moser,W., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogutu,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,K.,
Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Umanal,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 231758)
Worley,K.C.
Direct Submission
Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced gi:9966929.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project information
Center project name: UC
Center clone name: RP11-42B1

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224788 bases at least Q40
Consensus quality: 229074 bases at least Q30
Consensus quality: 230948 bases at least Q20
Estimated insert size: 227237; sum-of-contigs estimation
Estimated insert size: 317311; agarose-1p estimation
Quality coverage: 6.3x in Q20 bases; agarose-1p estimation
Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N. But the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 33241: contig of 33241 bp in length
33242 33341: gap of unknown length
33342 33341: contig of 23050 bp in length
56392 56491: gap of unknown length
56492 81323: contig of 24832 bp in length
81324 81423: gap of unknown length
81424 102538: contig of 21115 bp in length
102539 102638: gap of unknown length
102639 119710: contig of 17072 bp in length
119711 119810: gap of unknown length
119811 136913: contig of 17103 bp in length
136914 137013: gap of unknown length
137014 153285: contig of 16272 bp in length
153286 153385: gap of unknown length
153386 167987: contig of 14602 bp in length
167988 168087: gap of unknown length
168088 178731: contig of 10644 bp in length
178732 178831: gap of unknown length
178832 186641: contig of 7810 bp in length
186642 186741: gap of unknown length
186742 193215: contig of 6474 bp in length
193216 193315: gap of unknown length
193316 201310: contig of 7995 bp in length
201311 201410: gap of unknown length
201411 208647: contig of 7237 bp in length
208648 213802: gap of unknown length
213803 213902: contig of 5055 bp in length
213903 218049: gap of unknown length
218050 218149: gap of 4147 bp in length
218150 223316: gap of unknown length
223317 223416: contig of 5167 bp in length
223417 227389: gap of unknown length
227390 227489: contig of 3973 bp in length
227490 229032: gap of unknown length
229032 33241: contig of 1343 bp in length

```

* 229033 229132: gap of unknown length
* 229133 230651: contig of 1519 bp in length
* 230652 230751: gap of unknown length
* 230752 231758: contig of 1007 bp in length.
Location/Qualifiers
1. 231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"
BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

Query Match 96.8%; Score 30; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcggcgagcctcccccctcgtcgtc 30
|||||
Db 89335 CTCGGCGGCGCTCCCGCCCTTCGTGCTC 89306

RESULT 13
AF041472 4225 bp mRNA linear ROD 28-NOV-2001
LOCUS Mus musculus ataxin-2 (SCA2) mRNA, complete cds.
DEFINITION AF041472
ACCESSION AF041472.1 GI:3005019
VERSION AF041472.1 GI:3005019
KEYWORDS
SOURCE mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4225)
Nechiporuk,T.T., Huynh,D.P., Figueroa,K., Sahba,S., Nechiporuk,A.V.
and Pulst,S.M.
The mouse SCA2 gene: cDNA sequence, alternative splicing and
protein expression
Hum. Mol. Genet. 7 (8), 1301-1309 (1998)
968173
MEDLINE 98334550
PUBMED 2 (bases 1 to 4225)
Nechiporuk,T.T., Figueroa,K., Sahba,S., Nechiporuk,A.V. and
Pulst,S.M.
Direct Submission
TITLE Submitted (07-JAN-1998) Medicine/Neurology, Cedars-Sinai Medical
JOURNAL Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
AUTHORS
FEATURES
source
1. 4225
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="12q23.1"
1. 4225
/gene="SCA2"
27. 3884
/gene="SCA2"
/product="ataxin-2"
/protein_id="AAC09275.1"
/db_xref="GI:3005020"
/translation="MRSSTAAVQRPAAQDEPRRPGAWARSLPTARRRGKGVAV
YASAGPPRPGAPRGPSPPCASDCFSNCGASRPSRLIGVCGPPRPVVLL
ALAPATPARACPGVGRASPPRSGVSSARPGCGPACPEVYGLTMSLKPPQPP
APATGRKPGGGLISPGAPASAAVTSASVAPAPAPVASSSAAAGRGCGIGRGNS
SKGLPPTISFPGITYANVAKMHLITSVSKCEVQKNGITGVKRTSPKCDLVLD
AAHEKTESSSGKREIMESVLFKCSDFVVOFKTDSSTARDAFTDSALSAKVG
EKHEKLEPMADGELTASELELENDVSGMDPNDMFRNEENVGVSTYDSSLST
VPLENDSEELKREARANOIAEELISSAOYKARVLENDDESEERKYAVORNCSDR
EGHGPNTRDNKYITPCORNRREVLSMGSGROSSPRMGCGSMRPAASHTSDFNPA
GSDQRYVNGVGVMPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
HPSAHGSPAPVSTMPKRMSSSEGPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
SEMAAPPVARTSPAGTSSVSVGVPRLSPTKTHRPSRPOSSIGNSGSPGLASPOAG

```

```

ITPAEAVSPVPAASPTPASPNRALTPTSTIAKSDRIODORONSPAGSKENVKASET
SPSESKADNKGKSPVYSEHRKIDDLKTRKNDFRQPSSTSSNQLLSKNREBESK
DLIKRTERASANDSFIDSSSSNCTSGSKTNSPISMSLSNAEHRKGEVTSOGV
QTSPPACKQEKDREKKDPTQVRSSTLNPNAKFEFNRSFSQKPTTPTSPROAQ
PSPSPAGHQEPAPVYTOQVCEAPNMYEPVPSGVOPLYLIPMPMPVNOAKTVRAGK
VPNMPQORODHOHSTMMHAPASAGPPVATPSPAYSTOVAVSPQFNOPLVOHVP
YOSHPHYVSPVIOGNARMMAPPAHAQPELVSSAAGAEOTRAHACRILPYNKE
TSPSFYFASTGSLAQOYHAPNALHPHNPQPSATPTPGCOOSGSHGSPSPVOH
HOHQAAQALHLASPOOQSAITAGLAPTPSPSTPASNTQSPSSPPAAQOTVFTTHPS
HVQAPATTPPAHAHVQAHVQSGMVPSPHTAHAPMLMTTPPGPKAALQASALQPIP
VSTTAHFPMTHPSVQAHHQOL"
BASE COUNT 1007 a 1324 c 1042 g 851 t 1 others
ORIGIN

Query Match 86.5%; Score 26.8; DB 10; Length 4225;
Best Local Similarity 93.3%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcggcgagcctcccccctcgtcgtc 31
|||||
Db 295 TCTGCGGCGCTCCCGCCCTTCGTGCTG 324

RESULT 14
AF330031 303 bp DNA linear PRI 08-NOV-2001
LOCUS Macaca mulatta SCA2 gene, partial sequence.
DEFINITION AF330031
ACCESSION AF330031
VERSION AF330031.1 GI:12382833
KEYWORDS
SOURCE rhesus monkey.
ORGANISM Macaca mulatta.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 303)
Choudhury,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
JOURNAL 2 (bases 1 to 303)
Choudhury,S. and Brahmachari,S.K.
REFERENCE Direct Submission
TITLE Submitted (21-DEC-2000) Functional Genomics Unit, Center for
JOURNAL Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
source
1. 303
Location/Qualifiers
/organism="Macaca mulatta"
/db_xref="taxon:9544"
<1. .>303
/gene="SCA2"
/feature="Sinocebellar ataxia 2"
BASE COUNT 32 a 143 c 92 g 36 t
ORIGIN

Query Match 83.9%; Score 26; DB 9; Length 303;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ggggagcctcccccctcgtcgtc 30
|||||
Db 1 GCGGCGCTCCCGCCCTTCGTGCTC 26

RESULT 15
AF330028 390 bp DNA linear PRI 08-NOV-2001
LOCUS Pan troglodytes SCA2 gene, partial sequence.
DEFINITION

```

ACCESSION AF330028
VERSION AF330028.1 GI:12382830
KEYWORDS
SOURCE chimpanzee.
ORGANISM Pan troglodytes
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
AUTHORS 1 (bases 1 to 390)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
REFERENCE 2 (bases 1 to 390)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
source location/Qualifiers
1..390
/organism="Pan troglodytes"
/db_xref="taxon:9596"
repeat_region 1..390
/note="microsatellite"
/rpt_type-tandem
/rpt_unit-cag
<1..>390
gene /gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 48 a 183 c 110 g 49 t
ORIGIN

Query Match 83.9%; Score 26; DB 9; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gcgggctcccgcccttcgctgc 30
|||||
Db 1 GCGGGCTCCCGCCCTTCGCTC 26

Search completed: August 14, 2002, 21:48:16
Job time: 13514 sec

DR P-PSDB; AAW41370.
 XX Diagnosing spinocerebellar ataxia type II - by PCR and determining
 PT number of CAG repeat units
 XX
 XX Claim 1; Page 10; 23pp; Japanese.
 CC This sequence represents a fragment of the SCA2 gene. It can be used in
 CC the method of the invention for diagnosing spinocerebellar ataxia type
 CC II, by performing PCR on the test DNA using two primers hybridising to
 CC parts of the SCA2 gene sequence, and determining the number of CAG
 CC repeats in the amplified products. The method provides an easy means for
 CC the diagnosis of spinocerebellar ataxia type II.
 XX
 SQ Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

Query Match 100.0%; Score 31; DB 19; Length 355;
 Best local Similarity 100.0%; Pred. No. 0.037;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctccgcgggcctcccccctgcgtcgcg 31
 |||||||||||||||||||||||||||||
 Db 14g ctccgcgggcctcccccctgcgtcgcg 179

RESULT 2
 AAV06551
 ID AAV06551 standard; DNA; 516 BP.
 XX
 AC AAV06551;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE SCA2 gene fragment including CAG repeat region.
 XX
 XX SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT primer_bind complement (241..257)
 FT /tag= a
 FT /note= "primer SCA2-A binding site"
 FT primer_bind 349..366
 FT /tag= b
 FT /note= "primer SCA2-B binding site"
 FT exon 499..500
 FT /tag= c
 FT /note= "predicted splice site"
 FT repeat_region 267..332
 FT /tag= d
 FT /note= "CAG repeat region"
 FT repeat_unit 267..269
 FT /tag= e
 FT /note= "CAG repeat"
 FT repeat_unit 270..272
 FT /tag= f
 FT /note= "CAG repeat"
 FT repeat_unit 273..275
 FT /tag= g
 FT /note= "CAG repeat"
 FT repeat_unit 276..278
 FT /tag= h
 FT /note= "CAG repeat"
 FT repeat_unit 279..281
 FT /tag= i
 FT /note= "CAG repeat"
 FT repeat_unit 282..284
 FT /tag= j
 FT /note= "CAG repeat"
 FT repeat_unit 285..287

FT /tag= k
 FT /note= "CAG repeat"
 FT 291..293
 FT /tag= l
 FT /note= "CAG repeat"
 FT 294..296
 FT /tag= m
 FT /note= "CAG repeat"
 FT 297..299
 FT /tag= n
 FT /note= "CAG repeat"
 FT 300..302
 FT /tag= o
 FT /note= "CAG repeat"
 FT 306..308
 FT /tag= p
 FT /note= "CAG repeat"
 FT 309..311
 FT /tag= q
 FT /note= "CAG repeat"
 FT 312..314
 FT /tag= r
 FT /note= "CAG repeat"
 FT 315..317
 FT /tag= s
 FT /note= "CAG repeat"
 FT 318..320
 FT /tag= t
 FT /note= "CAG repeat"
 FT 321..323
 FT /tag= u
 FT /note= "CAG repeat"
 FT 324..326
 FT /tag= v
 FT /note= "CAG repeat"
 FT 327..329
 FT /tag= w
 FT /note= "CAG repeat"
 FT 330..332
 FT /tag= x
 FT /note= "CAG repeat"
 FT
 FT
 XX WO9742314-A1.
 XX PN
 XX 13-NOV-1997.
 XX PD
 XX 08-MAY-1997; 97WO-US07725.
 XX PF
 XX 08-OCT-1996; 96US-0727084.
 XX PR 08-MAY-1996; 96US-0017388.
 XX PR 19-JUL-1996; 96US-0022207.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst S;
 XX
 XX WPI; 1998-086523/08.
 XX
 PT Nucleic acids encoding human and mouse ataxin 2 - a product of the
 PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
 PT ataxia type 2
 XX
 XX
 PS Example 2; Page 51-52; 98pp; English.
 XX
 CC This genomic DNA in plasmid pL6512B includes a CAG repeat region
 CC from the novel human SCA2 gene (see AAV06552). It was identified
 CC following the construction of a bacterial artificial chromosome
 CC containing a pl artificial chromosome of the spinocerebellar
 CC ataxia 2 (SCA2) gene region and the identification of the SCA2
 CC gene from this contiguous map unit using a technique that screens
 CC for the presence of DNA trinucleotide repeats. The SCA2 locus is
 CC at 12q24.1. Ataxia type 2 can be diagnosed by detecting a genomic
 CC or transcribed mRNA sequence in an individual having an expanded

CC CAG repeat at a location corresponding to the CAG repeat region of
 CC the SCA2 gene. The presence of at least 13 CAG repeats above the
 CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
 CC primers (see AAV9640-41) amplifying at least this region are used
 CC for diagnosis. Also claimed are full-length ataxin-2 cDNAs for
 CC human and mouse (see AAV06552-53), kits for detecting mutations at
 CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
 CC useful for studying the physiological roles of SCA2 polypeptide
 CC (ataxin-2, see AAV33807-08) and its effect upon behaviour.

XX Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;

Query Match 100.0%; Score 31; DB 19; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.035;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcgcgggcctccccccttcgtcgtcg 31
 |||||
 DB 60 ctcgcgggcctccccccttcgtcgtcg 90

RESULT 3

AAV17229 standard; DNA; 623 BP.

AC AAV17229;

DT 29-JUN-1998 (first entry)

DE SCA2 gene fragment.

KW SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

OS Synthetic.

Key Location/Qualifiers

FT CDS 341..583 /tag- a

FT /note- "SCA2 protein fragment, no stop codon given"

XX WO9803679-A1.

XX 29-JAN-1998.

XX 18-JUL-1996; 96WO-JP01999.

XX 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sanpei K, Tsuji S;

XX WPI; 1998-120796/11.

XX P-PSDB; AAW41372.

XX Diagnosing spinocerebellar ataxis type II - by PCR and determining

XX number of CAG repeat units

XX Example 1; Page 11-12; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in

XX the method of the invention for diagnosing spinocerebellar ataxis type

XX II, by performing PCR on the test DNA using two primers hybridizing to

XX parts of the SCA2 gene sequence, and determining the number of CAG

XX repeats in the amplified products. The method provides an easy means for

XX the diagnosis of spinocerebellar ataxis type II.

SO Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 100.0%; Score 31; DB 19; Length 623;
 Best Local Similarity 100.0%; Pred. No. 0.035;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcgcgggcctccccccttcgtcgtcg 31
 |||||
 DB 149 ctcgcgggcctccccccttcgtcgtcg 179

RESULT 4

AAV78912 standard; cDNA; 4200 BP.

AC AAV78912;

DT 09-FEB-1998 (first entry)

DE Spinocerebellar ataxia gene SCA2.

KW Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;

KW repeat region; affinity; TATA binding protein; Kennedy disease;

KW transcription initiation factor; lymphoblastic cell line; schizophrenia;

KW Huntington's disease; dominant autosomal spinocerebellar ataxia;

KW X-linked spinocerebellar ataxia; familial spastic paraplegia;

KW dentatorubral-pallidolusian atrophy; bipolar affective disorder;

KW manic depressive psychosis; ss.

KW Homo sapiens.

Key Location/Qualifiers

FT CDS 3..2747 /tag- a

FT /product- SCA2 protein

FT /note- "this CDS contains a putative translational start

FT codon for the SCA2 protein at positions 243-245"

FT CDS 2594..3640 /tag- b

FT /note- "this second open reading frame may be derived

FT by a frameshift or by alternative splicing"

FT CDS 3..242 /tag- c

FT /note- "putative open reading frame which is in frame

FT with the putative translational start site of

FT the SCA2 open reading frame"

FT CDS 239..245 /tag- d

FT /note- "putative Kozak consensus signal"

FT repeat_region 258..323 /tag- e

FT /note- "encodes polyglutamine repeat region; contains

FT repeats of CAG with 2 CAA codons interspersed"

FT repeat_unit 258..260 /tag- f

FT /note- "CAG repeats"

FT misc_feature 1..3986 /tag- g

FT /note- "sequence contained in DAN1 clone"

FT misc_feature 3987..4200 /tag- h

FT /note- "derived from the EST's AAV92640, AAV90240 and

FT AA213574 from dbEST database"

FT misc_feature 4023..4029 /tag- i

FT /note- "region which differs in length between the

FT sequences of the EST clones AAV92640, AAV90240

FT and AA213574"

XX WO9717445-A1.

XX 15-MAY-1997.

XX PD 08-NOV-1996; 96WO-FR01773.

XX PR 10-NOV-1995; 95FR-0013576.

PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (INRM) INSERM INST NAT SANTE 6 RECH MEDICALE.
 XX
 PI Lutz Y, Mandel J, Tora L, Trotlier Y;
 DR WPI: 1997-281034/25.
 XX P-PSDB: AAW24800, AAW24801.
 XX
 PT Antibody 1C2 used for treating or preventing neuro-degenerative
 PT diseases - associated with proteins containing long poly:glutamine
 PT repeats, e.g. Huntington's disease
 XX
 PS Claim 21: Page 45-47: 69pp; French.
 XX
 CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
 CC treatment of neurodegenerative diseases associated with the presence
 CC of polyglutamine repeat regions. This Mab is already known for its
 CC affinity to the YAPTA binding protein (TBP) transcription initiation
 CC factor, especially at the amino acid sequence LEEQGRQDQDQ found at
 CC the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
 CC for polyglutamine repeats with a proportional affinity to the number
 CC of glutamine repeats. This affinity has been used to identify genes
 CC encoding proteins containing long polyglutamine repeats which are
 CC implicated in neurodegenerative diseases. A screen of an expression
 CC library, generated from a lymphoblastic cell line from a patient
 CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
 CC new sequences (AA78906-T78911) encoding polyglutamine repeats. Mab 1C2
 CC also isolated the complete SCA2 gene in clone DAN1 (sequence presented
 CC here). The sequence appears to contain 2 open reading frames (ORF) the
 CC second of which may be generated by an frameshift slippage or by an
 CC alternative splicing event. The first ORF also encodes a 22 amino acid
 CC polyglutamine repeat region near the N-terminus of the protein. Normal
 CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
 CC interspersed whereas the mutant sequence from patients with SCA
 CC contains at least 30, preferably 37-50 CAG repeats.
 CC Mab 1C2, active fragment of it or nucleic acids encoding it are
 CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
 CC X-linked spino-bulbar muscular atrophy (Kennedy disease),
 CC dentatorubral-pallidoluysal atrophy, dominant autosomal spinocerebellar
 CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
 CC depressive psychoses and schizophrenia.
 XX
 SQ Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

Query Match 100.0%; Score 31; DB 18; Length 4200;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgggcctcccccctctgctgcg 31
 |||||||||||||||||||||||||||||
 DB 51 ctcggcgggcctcccccctctgctgcg 81

RESULT 5
 AAV30270
 ID AAV30270 standard; DNA; 4367 BP.
 XX
 AC AAV30270;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
 XX
 KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
 KW CAG repeat; neurodegenerative disease; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 49..3990
 FT /*tag= a
 FT /product= "Spinocerebellar ataxia type 2 associated

FT repeat_region 544..612 protein"
 FT /*tag= b
 FT /note= "normal CAG repeat region; this is increased in
 FT repeat_unit 544..546 patients with SCA2"
 FT /*tag= c
 XX
 PN WO9818920-A1.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97MO-JP03946.
 XX
 PR 30-OCT-1996; 96JP-0304059.
 XX
 PA (SRUS-) SRL INC.
 XX
 PI Sanpei K, Tsuji S;
 XX
 DR WPI: 1998-272215/24.
 XX P-PSDB: AAW60213.
 XX
 PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
 PT - contain increased number of CAG repeat region compared to normal
 PT gene
 XX
 PS Claim 1; Pages 13-22; 38pp; Japanese.
 XX
 CC This represents the sequence of a gene causative of spinocerebellar
 CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
 CC with SCA2 has a tri-nucleotide (CAG) repeat region which in the
 CC expression product produces a polyglutamine sequence from Gln-166 to
 CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
 CC patients this number is increased to 35-100. Peptides encoded by nucleic
 CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
 CC gene, antibodies recognising the peptides and antisense nucleic acids
 CC hybridising with the nucleic acid fragments can be used for the
 CC investigation and diagnosis of SCA2. They can also be used for the
 CC treatment of SCA2 by antisense therapy or gene therapy.
 XX
 SQ Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

Query Match 100.0%; Score 31; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred. No. 0.03;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcggcgggcctcccccctctgctgcg 31
 |||||||||||||||||||||||||||||
 DB 337 ctcggcgggcctcccccctctgctgcg 367

RESULT 6
 AAV06552
 ID AAV06552 standard; CDNA; 4481 BP.
 XX
 AC AAV06552;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human SCA2 cDNA including CAG repeat region.
 XX
 KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 164..4101
 FT /*tag= a
 FT primer_bind complement (631..648)
 FT /*tag= b


```

FT /note= "primer SCA2-A binding site"
FT primer_bind 740..757
FT /tag= c
FT /note= "primer SCA2-B binding site"
FT primer_bind 1070..1091
FT /tag= d
FT /note= "primer SCA2-14B binding site"
FT exon 899..900
FT /tag= e
FT /note= "predicted splice site"
FT repeat_region 658..723
FT /tag= f
FT /note= "CAG repeat region"
FT repeat_unit 658..660
FT /tag= g
FT /note= "CAG repeat"
FT repeat_unit 661..663
FT /tag= h
FT /note= "CAG repeat"
FT repeat_unit 664..666
FT /tag= i
FT /note= "CAG repeat"
FT repeat_unit 667..669
FT /tag= j
FT /note= "CAG repeat"
FT repeat_unit 670..672
FT /tag= k
FT /note= "CAG repeat"
FT repeat_unit 673..675
FT /tag= l
FT /note= "CAG repeat"
FT repeat_unit 676..678
FT /tag= m
FT /note= "CAG repeat"
FT repeat_unit 679..681
FT /tag= n
FT /note= "CAG repeat"
FT repeat_unit 685..687
FT /tag= o
FT /note= "CAG repeat"
FT repeat_unit 688..690
FT /tag= p
FT /note= "CAG repeat"
FT repeat_unit 691..693
FT /tag= q
FT /note= "CAG repeat"
FT repeat_unit 694..696
FT /tag= r
FT /note= "CAG repeat"
FT repeat_unit 700..702
FT /tag= s
FT /note= "CAG repeat"
FT repeat_unit 703..705
FT /tag= t
FT /note= "CAG repeat"
FT repeat_unit 706..708
FT /tag= u
FT /note= "CAG repeat"
FT repeat_unit 709..711
FT /tag= v
FT /note= "CAG repeat"
FT repeat_unit 712..714
FT /tag= w
FT /note= "CAG repeat"
FT repeat_unit 715..717
FT /tag= x
FT /note= "CAG repeat"
FT repeat_unit 718..720
FT /tag= y
FT /note= "CAG repeat"
FT repeat_unit 721..723
FT /tag= z
FT /note= "CAG repeat"

```

```

XX PN MO9742314-A1.
XX PD 13-NOV-1997.
XX PF 08-MAY-1997; 97WO-US07725.
XX PR 08-OCT-1996; 96US-0727084.
XX PR 08-MAY-1996; 96US-0017388.
XX PR 19-JUL-1996; 96US-0022207.
XX PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX PI Pulst S;
XX PI MPI; 1998-086523/08.
XX DR P-PSDB; AAW33807.
XX DR
XX Nucleic acids encoding human and mouse ataxin 2 - a product of the
PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
PT ataxia type 2
PS
PS Claim 6; Page 52-58; 98pp; English.
XX
XX This cDNA sequence corresponds to a novel SCA2 gene encoding a human
CC spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
CC (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
CC human frontal cortex cDNA library in lambda Zapit were screened
CC with probes obtained by PCR amplification of plasmid AAP65122B (see
CC AAV06551). PCR products were used to screen the human adult frontal
CC cortex library, and 5' clones were obtained by RT-PCR of placental
CC mRNAs. Overlapping clones was used to generate the composite 4481
CC bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
CC or transcribed mRNA sequence in an individual having an expanded
CC CAG repeat at a location corresponding to the CAG repeat region of
CC the SCA2 gene. The presence of at least 13 CAG repeats above the
CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
CC Primers (see AAR99640-41) amplifying at least this region are used
CC for diagnosis. Also claimed are kits for detecting mutations at
CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
CC useful for studying the physiological roles of ataxin-2 and its
CC effect upon behaviour.
CC
CC
CC Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;
SQ

```

Query Match 100.0%; Score 31; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ctgcgcgggcctcccgcccttcgctcg 31
DB 451 ctgcgcgggcctcccgcccttcgctcg 481

```

RESULT 7
AA223428
ID AA223428 standard; DNA: 4481 BP.
XX
AC AA223428;
XX
XX 19-JAN-2000 (first entry)
DE Human SCA2 DNA.
XX
XX Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidoluystan atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic; ss.
XX
XX Homo sapiens.
OS

```
XX Key Location/Qualifiers
FH 163..4101
FT /*tag= a
FP /product= "SCA2"
XX
XX WO9945944-A1.
XX
XX 16-SEP-1999.
XX
XX 11-MAR-1999; 99WO-US05250.
XX
XX 12-MAR-1998; 98US-0041886.
XX
XX (BURN-) BURNHAM INSTR.
XX
XX Bredesen DE, Rabizadeh S;
XX
XX WPI; 1999-561617/47.
XX
XX P-PSDB; AAY33495.
XX
XX New proapoptotic dependence peptides, used to develop products for
XX treating, e.g., Alzheimer's disease.
XX
XX Disclosure; Page 130-135; 199pp; English.
XX
XX This invention describes novel pure proapoptotic dependence peptides
XX which comprise a sequence of an active dependence domain selected from
XX dependence polypeptides consisting of p75NMR, androgen receptor, DCC,
XX huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
XX SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
XX of inducing cell death and can be used to develop products to mediate or
XX inhibit apoptosis. The methods can be used for reducing the severity of
XX a proapoptotic dependence domain mediated pathological conditions e.g.
XX Huntington's disease, Alzheimer's disease, Kennedy's disease,
XX Spinocerebellar ataxia, dentatorubropallidolusian atrophy,
XX Machado-Joseph disease, stroke or head trauma. They can also be used for
XX reducing the severity of a pathological condition mediated by upregulated
XX cell proliferation or cell survival e.g. neoplastic, malignant,
XX autoimmune or fibrotic conditions. This sequence encodes the human
XX SCA2 polypeptide described in the method of the invention.
XX
XX Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 31; DB 20; Length 4481;
XX Best Local Similarity 100.0%; Pred. No. 0.03;
XX Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ctcgagcgagctccccccttcgtcgtcg 31
XX |||||||
DB 451 ctcgagcgagctccccccttcgtcgtcg 481
XX
XX RESULT 8
XX AAC56198/c
XX ID AAC56198 standard; DNA; 1008 BP.
XX
XX AAC56198;
XX
XX 25-JAN-2001 (first entry)
XX
XX Eucalyptus grandis transcription factor DNA sequence #329.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bzip; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Eucalyptus grandis.
XX
XX WO200053724-A2.
```

```
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide
XX
XX Claim 1; Page 131; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX sequence for one such transcription factor. The transcription factor may
XX be used to produce a plant having modified gene expression such as a
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX mahogany species or to modify the activity of a polypeptide in a plant.
XX The transcription factors of the present invention are members from the
XX following families of regulatory proteins: bzip, bzip family of G-box
XX binding factors, basic helix-loop-helix zipper,
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
XX and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX and MYB.
XX
XX Sequence 1008 BP; 175 A; 315 C; 331 G; 187 T; 0 other;
XX
XX
XX Query Match 69.0%; Score 21.4; DB 21; Length 1008;
XX Best Local Similarity 80.6%; Pred. No. 67;
XX Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 1 ctcgagcgagctccccccttcgtcgtcg 31
XX | ||||| || ||||| || ||
DB 340 CGCGCGGCGCCGCCGCCCTTCTTCCCG 310
XX
XX RESULT 9
XX AAK89967/c
XX ID AAK89967 standard; DNA; 726 BP.
XX
XX AAK89967;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human digestive system antigen genomic sequence SEQ ID NO: 3543.
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
XX
XX WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
```

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234927.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235464.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-502630/55.
XX
DR
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognostizing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3543; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 726 BP; 139 A; 209 C; 200 G; 178 T; 0 other;

Query Match 66.5%; Score 20.6; DB 22; Length 726;
Best Local Similarity 85.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctccgagcgcctcccccgcctcgc 27
||| ||||| ||||| ||||| |||||
Db 500 CTCTCGGGCTCCCTGCCCGTGTCTC 474

RESULT 10
AAL34906
ID AAL34906 standard; cDNA; 328 BP.
XX
AC AAL34906;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 248.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01338.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231043.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-02441826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.

08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-451937/48.
XX P-PSDB; ABB03324.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders related to the musculoskeletal system including
XX musculoskeletal cancers and also for testing and detection e.g.
XX diagnosis -
XX
XX Claim 1: SEQ ID NO 248; 781bp + Sequence Listing: English.
XX
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
XX (ABB03087-ABB04109) associated with the musculoskeletal system useful
XX for preventing, treating or ameliorating medical conditions e.g. by
XX protein or gene therapy. The genes are isolated from a range of human
XX tissues disclosed in the specification. The nucleic acids, proteins,
XX antibodies and (ant)agonists are useful in the diagnosis, treatment
XX and prevention of: (a) cancer, e.g. breast and ovarian cancer and
XX other cancers of the adrenal gland, bone, bone marrow, breast,
XX gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
XX (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
XX healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
XX and (f) infectious diseases such as viral, bacterial, fungal and
XX parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 328 BP; 57 A; 93 C; 113 G; 65 T; 0 other;

Query Match 65.8%; Score 20.4; DB 22; Length 328;

Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 ctcgcggcgccctccccccttcgctgc 30
||||| ||||| ||||| || ||
Db 88 ctcgcgcggcgccctccccccttcgctgc 117
RESULT 11
ABLI3077
ID ABLI3077 standard; cDNA; 3592 BP.
XX
XX ABLI3077;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33713.
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33713.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX P-PSDB; ABB68974.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 33713; 21bp + Sequence Listing: English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3592 BP; 968 A; 904 C; 899 G; 821 T; 0 other;
XX
XX Query Match 65.8%; Score 20.4; DB 23; Length 3592;
XX Best Local Similarity 80.0%; Pred. No. 1.3e+02;
XX Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 ctcgcggcgccctccccccttcgctgc 30
||||| ||||| ||||| || ||
Db 373 cgcgcggcgccctccccccttcgctgc 402
RESULT 12
ABLI3076/C
ID ABLI3076 standard; cDNA; 17500 BP.
XX

AC ABL13076;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33710.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EM;
XX
DR WPI: 2001-656860/75.
XX
DR P-PSDB: ABB68973.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 33710; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABH57737-ABH72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 17500 BP; 5162 A; 3589 C; 3637 G; 5112 T; 0 other;

Query Match 65.8%; Score 20.4; DB 23; Length 17500;
Best Local Similarity 80.0%; Pred. No. 1.2e-02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 ctcggggggcccccgcctcgcgcgc 30
| ||||| ||||| ||||| ||||| |||||
Db 15742 CGCGGCGCCCTCCGCTCCGCTC 15713

RESULT 13
ID AAL36153 standard; DNA; 796 BP.
XX
AC AAL36153;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2518.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
XX
PR 24-FEB-2000; 2000US-0184664.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PR 16-MAR-2000; 2000US-0189874.
XX
PR 17-MAR-2000; 2000US-0190076.
XX
PR 18-APR-2000; 2000US-0198123.
XX
PR 19-MAY-2000; 2000US-0205515.
XX
PR 07-JUN-2000; 2000US-0209467.
XX
PR 28-JUN-2000; 2000US-0214886.
XX
PR 30-JUN-2000; 2000US-0215135.
XX
PR 07-JUL-2000; 2000US-0216647.
XX
PR 07-JUL-2000; 2000US-0216880.
XX
PR 11-JUL-2000; 2000US-0217487.
XX
PR 11-JUL-2000; 2000US-0217496.
XX
PR 14-JUL-2000; 2000US-0218290.
XX
PR 26-JUL-2000; 2000US-0220963.
XX
PR 14-AUG-2000; 2000US-0224518.
XX
PR 14-AUG-2000; 2000US-0224519.
XX
PR 14-AUG-2000; 2000US-0225213.
XX
PR 14-AUG-2000; 2000US-0225214.
XX
PR 14-AUG-2000; 2000US-0225267.
XX
PR 14-AUG-2000; 2000US-0225268.
XX
PR 14-AUG-2000; 2000US-0225270.
XX
PR 14-AUG-2000; 2000US-0225447.
XX
PR 14-AUG-2000; 2000US-0225757.
XX
PR 14-AUG-2000; 2000US-0225758.
XX
PR 14-AUG-2000; 2000US-0225759.
XX
PR 18-AUG-2000; 2000US-0226279.
XX
PR 22-AUG-2000; 2000US-0226681.
XX
PR 22-AUG-2000; 2000US-0226868.
XX
PR 22-AUG-2000; 2000US-0227182.
XX
PR 30-AUG-2000; 2000US-0227029.
XX
PR 01-SEP-2000; 2000US-0228924.
XX
PR 01-SEP-2000; 2000US-0229287.
XX
PR 01-SEP-2000; 2000US-0229343.
XX
PR 01-SEP-2000; 2000US-0229344.
XX
PR 01-SEP-2000; 2000US-0229345.
XX
PR 05-SEP-2000; 2000US-0229509.
XX
PR 05-SEP-2000; 2000US-0229513.
XX
PR 06-SEP-2000; 2000US-0230437.
XX
PR 06-SEP-2000; 2000US-0230438.
XX
PR 08-SEP-2000; 2000US-0231242.
XX
PR 08-SEP-2000; 2000US-0231243.
XX
PR 08-SEP-2000; 2000US-0231244.
XX
PR 08-SEP-2000; 2000US-0231413.
XX
PR 08-SEP-2000; 2000US-0231414.
XX
PR 08-SEP-2000; 2000US-0232080.
XX
PR 08-SEP-2000; 2000US-0232081.
XX
PR 12-SEP-2000; 2000US-0231968.
XX
PR 14-SEP-2000; 2000US-0232397.
XX
PR 14-SEP-2000; 2000US-0232398.
XX
PR 14-SEP-2000; 2000US-0232399.
XX
PR 14-SEP-2000; 2000US-0232400.
XX
PR 14-SEP-2000; 2000US-0232401.
XX
PR 14-SEP-2000; 2000US-0233063.
XX
PR 14-SEP-2000; 2000US-0233064.
XX
PR 14-SEP-2000; 2000US-0233065.
XX
PR 21-SEP-2000; 2000US-0234223.
XX
PR 21-SEP-2000; 2000US-0234274.
XX
PR 25-SEP-2000; 2000US-0234997.
XX
PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0235678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
XX
PS Example 2; SEQ ID NO 2518; 781pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL36154-AU37666) and proteins
CC (AB03087-AB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases such as cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 796 BP; 184 A; 233 C; 219 G; 160 T; 0 other;
XX
Query Match 65.2%; Score 20.2; DB 22; Length 796;
Best Local Similarity 88.0%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 ctcgcgcgcgcctcccccctctcg 25
Db 532 ctcgcgcgcgcctcccccctctcg 556
RESULT 14
AAL36154
ID AAL36154 standard; DNA; 2930 BP.
XX
AC AAL36154;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2519.
XX
KW Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO20015367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
XX

XX The invention relates to novel genes (AA134669-AA137666) and proteins
CC (AB03087-AB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC anemias, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 2930 BP; 660 A; 841 C; 840 G; 589 T; 0 other;

Query Match 65.2%; Score 20.2; DB 22; Length 2930;
Best Local Similarity 88.0%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ctggcgagcctccgcccttcg 25
DB 1234 ctgcgcgcctccgcccttcg 1258
||||| ||||||| ||||||| ||

RESULT 15

AA566114/c

ID AA566114 standard; cDNA; 1101 BP.

XX AA566114;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #1918.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;

KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABC01927.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 1; SEQ ID NO 1918; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 1101 BP; 224 A; 323 C; 359 G; 194 T; 1 other;

Query Match 64.5%; Score 20; DB 23; Length 1101;
Best Local Similarity 82.1%; Pred. No. 2e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ctggcgagcctccgcccttcg 28
DB 674 CCCGCCGCCCTTCGCGCCCTTCG 647
||||| ||||||| ||||||| ||

Search completed: August 14, 2002, 22:06:34
Job time: 11689 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:51:15 ; Search time 203.42 seconds

(without alignments)
37.433 Million cell updates/sec

Title: US-09-707-919-4

Sequence: 1 ctgagcgagcctcccccctctgctgctg 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
4: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
5: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
6: /cgn2_6/pdata/1/lna/5B.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	355	4	US-09-043-303-1
2	31	100.0	623	4	US-09-043-303-5
3	31	100.0	4481	4	US-09-041-886-18
4	19	61.3	1743	3	US-09-032-365A-18
5	19	61.3	2575	4	US-09-077-354B-1
6	19	61.3	10380	4	US-09-077-354B-3
7	18.8	60.6	43804	4	US-09-171-461-1
8	18.8	60.6	4403765	4	US-09-103-840A-2
9	18.6	60.0	1962	4	US-08-791-115B-3
10	18.4	59.4	1200	1	US-08-356-397-10
11	18.4	59.4	2936	2	US-08-714-677-10
12	18.4	59.4	2936	2	US-08-393-540-10
13	18.4	59.4	2936	2	US-08-714-537-10
14	18.4	59.4	3196	2	US-09-096-982-4
15	18.4	59.4	3196	2	US-08-653-650A-4
16	18.4	59.4	4411529	4	US-09-103-840A-1
17	18.2	58.7	884	2	US-08-901-200A-11
18	18.2	58.7	884	3	US-09-219-391-11
19	18.2	58.7	1100	2	US-08-776-210-4
20	18.2	58.7	1144	1	US-08-014-943A-1
21	18.2	58.7	1144	1	US-08-486-421-2
22	18.2	58.7	1144	1	US-08-470-911-2
23	18.2	58.7	1144	1	US-08-486-809-2
24	18.2	58.7	1257	2	US-08-776-210-2
25	18.2	58.7	1723	1	US-07-841-646-28
26	18.2	58.7	1723	1	US-07-901-703-10
27	18.2	58.7	1723	1	US-08-147-023-28

c 28	18.2	58.7	1723	1	US-08-206-86A-3	Sequence 3, Appl1
c 29	18.2	58.7	1723	1	US-08-278-729A-20	Sequence 20, Appl1
c 30	18.2	58.7	1723	1	US-08-480-528A-7	Sequence 7, Appl1
c 31	18.2	58.7	1723	1	US-08-479-666-7	Sequence 7, Appl1
c 32	18.2	58.7	1723	1	US-08-155-343A-20	Sequence 20, Appl1
c 33	18.2	58.7	1723	1	US-08-406-672-20	Sequence 20, Appl1
c 34	18.2	58.7	1723	1	US-08-643-563A-20	Sequence 20, Appl1
c 35	18.2	58.7	1723	1	US-08-447-570A-28	Sequence 28, Appl1
c 36	18.2	58.7	1723	1	US-08-643-763A-20	Sequence 20, Appl1
c 37	18.2	58.7	1723	1	US-08-462-623-20	Sequence 20, Appl1
c 38	18.2	58.7	1723	1	US-08-451-953A-20	Sequence 20, Appl1
c 39	18.2	58.7	1723	2	US-08-459-346-5	Sequence 5, Appl1
c 40	18.2	58.7	1723	2	US-08-445-468A-20	Sequence 20, Appl1
c 41	18.2	58.7	1723	2	US-08-901-200A-7	Sequence 7, Appl1
c 42	18.2	58.7	1723	2	US-08-449-700-28	Sequence 28, Appl1
c 43	18.2	58.7	1723	2	US-08-449-699A-28	Sequence 28, Appl1
c 44	18.2	58.7	1723	2	US-08-461-397A-20	Sequence 20, Appl1
c 45	18.2	58.7	1723	2	US-08-912-088-20	Sequence 20, Appl1

ALIGNMENTS

```
RESULT 1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
; US-09-043-303-1
```

Query Match 100.0%; Score 31; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ctgagcgagcctcccccctctgctgctg 31
Db 149 ctgagcgagcctcccccctctgctgctg 179
```

```
RESULT 2
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 623
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (341)..(583)
FEATURE:
OTHER INFORMATION: Tsp-2
US-09-043-303-5

Query Match 100.0%; Score 31; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcggcgagctcccccgccttcgtctg 31
Db 149 ctcggcgagctcccccgccttcgtctg 179

RESULT 3
US-09-041-886-18
Sequence 18, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
ATTORNEY: Rabizadeh, Sharoz
TITLE OF INVENTION: Protoproctic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

Query Match 100.0%; Score 31; DB 4; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.0041;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctcggcgagctcccccgccttcgtctg 31

Db 451 CTCGGCGAGCTCCCCGCCCTTCGTCTG 481

RESULT 4
US-09-032-365A-18/c
Sequence 18, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Mishina, Patsy
APPLICANT: Naggart, Juergen
APPLICANT: No. 6114502en-Trauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2C1P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1743 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-032-365A-18

Query Match 61.3%; Score 19; DB 3; Length 1743;
Best Local Similarity 81.5%; Pred. No. 64;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 gccggagctcccccgccttcgtctg 31
Db 236 GCGCTCTCCCGCGCCTTGTCGCG 210

RESULT 5
US-09-077-354B-1
Sequence 1, Application US/09077354B
Patent No. 6255096
GENERAL INFORMATION:
APPLICANT: HOPKINS, JOHN JOSEPH; SCOTT, HAMISH STEELE;
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART
TITLE OF INVENTION: SYNTHETIC MAMMALIAN
TITLE OF INVENTION: \N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SA
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

```

ADDRESS: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 GARDEN CITY PLAZA
CITY: GARDEN CITY
STATE: NEW YORK
COUNTRY: UNITED STATES
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/077,354B
FILING DATE: 22-APRIL-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00747
FILING DATE: 22-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: POKALSKY, ANN R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 12416
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516 742 4343
TELEFAX: 516 742 4366
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Peripheral Blood
CELL TYPE: Leukocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 102..2330
US-09-077-354B-1

Query Match
Best Local Similarity 61.3%; Score 19; DB 4; Length 2575;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 gcgggctcccgccctgctgctg 31
    ||||| |||| |||| | |||||
Db 53 GCGGGCGCCCAACCCCTGCGCTCG 79

RESULT 6
US-09-077-354B-3
Sequence 3, Application US/09077354B
Patent No. 6235096
GENERAL INFORMATION:
APPLICANT: HOWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART
TITLE OF INVENTION: SYNTHETIC MAMMALIAN
TITLE OF INVENTION: '-N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 GARDEN CITY PLAZA
CITY: GARDEN CITY
STATE: NEW YORK
COUNTRY: UNITED STATES
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/09/077,354B
FILING DATE: 22-APRIL-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00747
FILING DATE: 22-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: POKALSKY, ANN R.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 12416
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516 742 4343
TELEFAX: 516 742 4366
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 10380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Chromosome 17
FEATURE:
NAME/KEY: exon 1
LOCATION: 990..1372
FEATURE:
NAME/KEY: exon 2
LOCATION: 2115..2262
FEATURE:
NAME/KEY: exon 3
LOCATION: 3056..3202
FEATURE:
NAME/KEY: exon 4
LOCATION: 3387..3472
FEATURE:
NAME/KEY: exon 5
LOCATION: 5667..5923
FEATURE:
NAME/KEY: exon 6
LOCATION: 7745..8955
US-09-077-354B-3

Query Match
Best Local Similarity 61.3%; Score 19; DB 4; Length 10380;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 gcgggctcccgccctgctgctg 31
    ||||| |||| |||| | |||||
Db 941 GCGGGCGCCCAACCCCTGCGCTCG 967

RESULT 7
US-09-171-461-1/C
Sequence 1, Application US/09171461
Patent No. 6335016
GENERAL INFORMATION:
APPLICANT: Baker, Adam
APPLICANT: Cotten, Matthew
APPLICANT: Chioocca, Susanna
APPLICANT: Kurzbauer, Robert
APPLICANT: Schaffner, Gotthold
TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
FILE REFERENCE: 0652,1800000
CURRENT APPLICATION NUMBER: US/09/171,461
CURRENT FILING DATE: 1999-01-12
EARLIER APPLICATION NUMBER: PCT/EP97/01944
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 43804
```

```

1 TYPE: DNA
2 ORGANISM: CELO Virus
3 FEATURE:
4 NAME/KEY: gene
5 LOCATION: (12193)..(15043)
6 OTHER INFORMATION: /gene: L1
7 FEATURE:
8 NAME/KEY: misc_feature
9 LOCATION: (15080)
10 OTHER INFORMATION: /note= L2 region penton base splice acceptor site
11 FEATURE:
12 NAME/KEY: gene
13 LOCATION: (15110)..(17495)
14 OTHER INFORMATION: /gene: L2
15 FEATURE:
16 NAME/KEY: polyA_site
17 LOCATION: (17526)
18 FEATURE:
19 NAME/KEY: gene
20 LOCATION: (17559)..(21754)
21 OTHER INFORMATION: /gene: L3
22 FEATURE:
23 NAME/KEY: misc_feature
24 LOCATION: (18261)
25 OTHER INFORMATION: /gene: L3 /note= hexon splice acceptor site
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: (21102)
29 OTHER INFORMATION: /gene: L3 /note= protease splice acceptor site
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: (21123)
33 OTHER INFORMATION: /gene: L3 /note= protease splice acceptor site
34 FEATURE:
35 NAME/KEY: polyA_site
36 LOCATION: (21767)
37 FEATURE:
38 NAME/KEY: polyA_site
39 LOCATION: (21824)
40 FEATURE:
41 NAME/KEY: polyA_site
42 LOCATION: (21836)
43 FEATURE:
44 NAME/KEY: polyA_site
45 LOCATION: (21882)
46 FEATURE:
47 NAME/KEY: misc_feature
48 LOCATION: (22608)
49 OTHER INFORMATION: /note= 100K splice acceptor site
50 FEATURE:
51 NAME/KEY: misc_feature
52 LOCATION: (22649)
53 OTHER INFORMATION: /note= 100K splice acceptor site
54 FEATURE:
55 NAME/KEY: gene
56 LOCATION: (23680)..(27886)
57 OTHER INFORMATION: /gene: L4
58 FEATURE:
59 NAME/KEY: polyA_site
60 LOCATION: (27920)
61 FEATURE:
62 NAME/KEY: misc_feature
63 LOCATION: (28315)
64 OTHER INFORMATION: /note= fibre splice acceptor site
65 FEATURE:
66 NAME/KEY: misc_feature
67 LOCATION: (28341)
68 OTHER INFORMATION: / note= fibre splice acceptor site
69 FEATURE:
70 NAME/KEY: gene
71 LOCATION: (28363)..(31768)
72 OTHER INFORMATION: /gene: L5
73 FEATURE:

```

```

? NAME/KEY: misc_feature
? LOCATION: (30511)
? OTHER INFORMATION: /gene: L5 /note- fibre splice acceptor site
? FEATURE:
? NAME/KEY: polyA_site
? LOCATION: (31770)
? US-03-171-461-1

```

Query Match	60.6%	Score 18.8	DB 4	Length 43804
Best Local Similarity	76.7%	Pred. No. 66		
Matches 23	Conservative	0	Mismatches	7
			Indels	0
			Gaps	0
QY	2	tcgagcgagctcccccgcctctgcgcgcg	31	
Db	17188	TCGCGGCGCTCCAAAGCGCGCTAGTCGCG	17159	

```

RESULT      8
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequ
US-09-103-840A-2

```

	Query Match	60.6%	Score 18.8:	DB 4;	Length 440365;
	Best Local Similarity	76.7%;	Pred No. 35;		
	Matches 23; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;
OY	2 tcgcgagcctcccccgcaccttcgtcgatc	31			
Db 829571	tTGGCGGtGTGACCGCTCCATTGTTCGTG	829542			

RESULT 9
 US-08-791-115B-3
 : Sequence 3, Application US/08791115B
 : Patent No. 6262242
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Steck, Peter
 : APPLICANT: Perhouse, Mark A.
 : APPLICANT: Jasser, Samir
 : APPLICANT: Yung, W.K. Alfred
 : APPLICANT: Tavilgjan, Sean V.
 :
 : TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10023.3
 :
 : NUMBER OF SEQUENCES: 27
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Rothwell, Flg9, Ernst & Kurtz, P.C.,
 : STREET: 555 Thirteenth Street, N.W., Suite 701-E
 : CITY: Washington
 : STATE: DC
 :
 : COUNTRY: USA
 : ZIP: 22204
 :

ORIGIN

Query Match 100.0%; Score 31; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcgagcgagcctcccgccctgcgtcg 31
|||||
Db 98 CTCGGCGGCGCTCCCGCCCTGCTGCTG 128

RESULT 2
B1547486 500 bp mRNA linear EST 05-SEP-2001
LOCUS 603191091F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262335 5',
DEFINITION mRNA sequence.
B1547486
ACCESSION B1547486.1 GI:15434798
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 500)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11661 row: e column: 24
High quality sequence stop: 485.
Location/Qualifiers
1. 500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262335"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 57 a 222 c 150 g 71 t

ORIGIN

Query Match 100.0%; Score 31; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcgagcgagcctcccgccctgcgtcg 31
|||||
Db 101 CTCGGCGGCGCTCCCGCCCTGCTGCTG 131

RESULT 3
B1545214 1100 bp mRNA linear EST 05-FEB-2002
LOCUS BM455214

DEFINITION AGENCOURT_6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
5', mRNA sequence.
ACCESSION BM455214
KEYWORDS BM455214.1 GI:18504254
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1100)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12134 row: k column: 12
High quality sequence stop: 623.
Location/Qualifiers
1. 1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 240 a 329 c 306 g 219 t 6 others

ORIGIN

Query Match 96.8%; Score 30; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcgagcgagcctcccgccctgcgtcg 30
|||||
Db 72 CTCGGCGGCGCTCCCGCCCTGCTGCTG 101

RESULT 4
BE457923 364 bp mRNA linear EST 26-JUL-2000
LOCUS BE457923
DEFINITION us99c12.x1 Soares_thymus_2NBMT Mus musculus cDNA clone
IMAGE:3326518 3' similar to TR:070305 070305 SPINOCEREBELLAR AVAXIA
2 HOMOLOG ;, mRNA sequence.
ACCESSION BE457923
VERSION BE457923.1 GI:9480561
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 364)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMGC Consortium (info@image.llnl.gov) for further information.
MGI:1070682

FEATURES
source Possible reversed clone: polyt not found.
Location/Qualifiers
1. 364

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="3326518"
/clone_lib="Soares_thymus_2NBMt"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 51 a 126 c 173 g 14 t

ORIGIN

Query Match 86.5%; Score 26.8; DB 10; Length 364;
Best Local Similarity 93.3%; Pred. No. 1.2e+02;
Matches 28; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tcggcgagccctcccgccctcgctcg 31

Db 346 TCTCGGGGCTCCCGCCCTTCGTCGTG 317

RESULT 5
BI478400 343 bp mRNA linear EST 27-AUG-2001
LOCUS BI478400
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION BI478400
VERSION BI478400.1 GI:15312818
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 343)

REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949065 row: D column: 08.
Location/Qualifiers

FEATURES
source

1. 343
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:

BASE COUNT 34 a 131 c 138 g 40 t

ORIGIN

Query Match 72.9%; Score 22.6; DB 10; Length 343;
Best Local Similarity 86.2%; Pred. No. 1.9e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cggcgagccctcccgccctcgctcg 31

Db 126 CGCGGGGCGCCCGCGCGCTCGTCGTCG 154

RESULT 6
AO747830 1030 bp DNA linear GSS 19-JUL-1999
LOCUS AO747830
DEFINITION HS_5537_A1_F03-SP6 RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1113 Col=5 Row=K, DNA sequence.
ACCESSION AO747830
VERSION AO747830.1 GI:5534988
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1030)

REFERENCE
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,D., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
MEDLINE Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
CONTACT: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 1030.
Location/Qualifiers

FEATURES
source

1. 1030
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="Plate=1113 Col=5 Row=K"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3 6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the

BASE COUNT 268 a 296 c 402 g 50 t 14 others
ORIGIN

Query Match 72.9%; Score 22.6; DB 12; Length 1030;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 ctgcgagcctccccccttcgtcgcg 31
||||| ||||||| ||||| |||||
Db 434 TCGCGGNGCTCCCGCCCTTCGCGCG 405

RESULT 7
BF586264 446 bp mRNA linear EST 12-DEC-2000
LOCUS
DEFINITION FMI_27_G04.b1_A003 Floral-Induced Meristem 1 (FMI) Sorghum
PROPIONQUUM cDNA, mRNA sequence.
ACCESSION BF586264
VERSION BF586264.1 GI:11678588
KEYWORDS
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 446)
Cordonier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 410
POLYA-No.

FEATURES
source

1..446 Location/Qualifiers
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FMI)"
/note="Organ: Floral-Induced meristems; Vector:
pbluescript II from lambda Zap II; Site 1: XhoI; Site 2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested. The
library was made from poly-A RNA in the cloning vector
lambda Zap II. Clones to be sequenced were prepared by
mass excision."

BASE COUNT 59 a 162 c 147 g 78 t
ORIGIN

Query Match 69.0%; Score 21.4; DB 10; Length 446;
Best Local Similarity 80.6%; Pred. No. 4.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 ctgcgagcctccccccttcgtcgcg 31
||||| ||||||| ||||| |||||
Db 310 CGCGCGGCGCGCCCGCCCTTCGCTGCTCG 340

RESULT 8
Bj177625/c 512 bp mRNA linear EST 24-JAN-2002
LOCUS
DEFINITION Bj177625 normalized full length cDNA library, chloronemata,
caulonemata and malformed buds Physcomitrella patens subsp. patens
cDNA clone pphb20j04 5', mRNA sequence.

ACCESSION Bj177625
VERSION Bj177625.1 GI:18345582
KEYWORDS
SOURCE
ORGANISM

Physcomitrella patens subsp. patens.
Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 512)
Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
,M.

TITLE Comparison of the moss Physcomitrella patens genome with flowering
plants genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genetics.nig.ac.jp

A backbone of the vector is pbluescript II, that was in vivo
excised from a modified lps phage vector (Mo bl rec, Germany). XhoI
digested-5' end of cDNA is ligated to SalI site of the vector, and
the BamHI digested-3' end including poly-A tail is ligated to BamHI
site of the vector. cDNA insert could be amplified with
conventional T7 and T3 primers. This normalized full-length cDNA
library was generated basically according to the method described
in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
Protonemata were blended by the POLYTRON, and then cultivated on
the BCD medium containing 0.5uM BA (benzylaminopurine) for 8 to 13
days under the continuous light.

FEATURES
source

1..512 Location/Qualifiers
/organism="Physcomitrella patens subsp. patens"
/db_xref="taxon:145481"
/clone_lib="pphb20j04"
/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and malformed buds"
/tissue_type="mixture of chloronemata, caulonemata and
malformed buds"

BASE COUNT 127 a 109 c 170 g 106 t
ORIGIN

Query Match 69.0%; Score 21.4; DB 10; Length 512;
Best Local Similarity 80.6%; Pred. No. 4.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 ctgcgagcctccccccttcgtcgcg 31
||||| ||||||| ||||| |||||
Db 255 CTCAGCGGCGCTCCCGCCCTTCGCTGCTCG 225

RESULT 9
BG817511 529 bp mRNA linear EST 22-MAY-2001
LOCUS
DEFINITION FMI_76_H05.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION BG817511
VERSION BG817511.1 GI:14188491
KEYWORDS
SOURCE
ORGANISM

Sorghum bicolor
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

```

REFERENCE 1 (bases 1 to 529)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingie,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 410
POLYA-No.

FEATURES
    source
        1..529
            /organism="Sorghum bicolor"
            /db_xref="taxon:4558"
            /clone_lib="Embryo 1 (EM1)"
            /note="Organ: Embryos germinated for 24 hr: Vector:
            pBluescript II from Lambda Zap II; Site.1: XhoI; Site.2:
            EcoRI; The library was made from polyA RNA in the cloning
            vector Lambda Zap II. Clones to be sequenced were
            prepared by mass excision."
            prepared by mass excision."

BASE COUNT      75 a      191 c      173 g      90 t
ORIGIN
Query Match      69.0%; Score 21.4; DB 10; Length 529;
Best Local Similarity 80.6%; Pred. No. 4.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy      1 ctcggcggggccctcccccgcctcgtcgctg 31
      | | | | | | | | | | | | | | | | | | |
Db      326 GCGCGCGGCGCGCGCCGCGCCGCTGCTGTCG 356

RESULT 10
BG356187      563 bp      mRNA      linear      EST 06-MAR-2001
LOCUS      BG356187
DEFINITION      EML_22.E07.bl_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
ACCESSION      BG356187
VERSION      BG356187.1 GI:13238173
KEYWORDS
SOURCE
    organism.
    Sorghum bicolor
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
    clade; Panicoideae; Andropogoneae; Sorghum.
    1 (bases 1 to 563)
    Reid,S.P., Cordonnier-Pratt,M.-M., Gingie,A. and Pratt,L.H.
    An EST database from Sorghum: developing embryos
    Unpublished (2000)
    Contact: Cordonnier-Pratt MM
    Department of Botany
    The University of Georgia
    Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
    Tel: 706 542 1860
    Fax: 706 542 1805
    Email: mmpratt@uga.edu
    Sequences have been trimmed to exclude PolyA, vector and regions
    below Phred quality 16. The threshold for highest quality sequence
    is 20.
    Seq primer: JEN REV
    High quality sequence stop: 503
    POLYA-No.

FEATURES
    source
        1..563
            /organism="Sorghum bicolor"
            location/Qualifiers

```

```

/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; The library was made from poly(A) RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT      76 a      208 c      188 g      90 t      1 others
ORIGIN

Query Match      69.0%; Score 21.4; DB 10; Length 563;
Best Local Similarity 80.6%; Pred. No. 4.2e+03;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy      1 ctcgagcgagctcccgcccttcgtcgtc 31
      1 | | | | | | | | | | | | | | | | | | | |
Db      313 CcCCCGGGGGGGCCCGCCCGCCTCGTCGCG 343

RESULT  11
LOCUS   BF631132
DEFINITION  BF631132      865 bp      mRNA      linear      EST 22-OCT-2001
          HVSME00015B02f Hordeum vulgare seedling shoot EST library
          HVCDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone
          HVSME00015B02f, mRNA sequence.
ACCESSION  BF631132
VERSION    BF631132.2  GI:13091913
KEYWORDS   EST.
SOURCE     barley.
ORGANISM   Hordeum vulgare
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
            ; Triticeae; Hordeum.
            1 (bases 1 to 865)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Fritsch,D., Yu
,R.D., Henry,D., Palmer,M., Rambo,T., Simons,J., Choi,D.W., Fenton
,R.D., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex drought-stressed seedling shoot cDNA
library
Unpublished (2001)
On Dec 19, 2000 this sequence version replaced gi:11895290.
COMMENT   Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hg bases - 164
            Seq primer: AATTAACTTCACGTAAAGCG
            High quality sequence stop: 186.
            Location/Qualifiers
                1..865
                /organism="Hordeum vulgare"
                /cultivar="Morex"
                /db_xref="taxon:4513"
                /clone="HVSME00015B02f"
                /clone_lib="Hordeum vulgare seedling shoot EST library
                HVCNA0002 (Dehydration stress)"
                /tissue-type="Seedling shoot"
                /lab_host="TJc121"
                /note="Vector: lambdaZap; Site_1: EcoRI; Site_2: XhoI.
                Seeds were surface sterilized then germinated under axenic
                conditions in the dark at room temperature on filter paper
                with water, nystatin and cefotaxime in covered
                crystallization dishes. Five-day old seedlings were
                incubated at 90% RH for 24 hr. Shoots were then harvested,
                total RNA was prepared, poly(A) RNA was purified, one
                primary unamplified cDNA library was made, 600000 pfu were
                in vivo excised to give pBluescript SK(-) cDNA phagemids.
                These steps were performed in the TJ Close laboratory at

```


1 (bases 1 to 1278)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC,
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Cell: 301-495-4700 ext. 1111
 Fax: 301-495-4700
 E-mail: strausberg@mail.nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.jnl.gov>
Plate: LILCH100 row: 1 column: 18
High quality sequence stop: 554.

Location/Qualifiers
1. .1278

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4277921"
 /clone_lib="NIH_MGC.56"
 /clone_type="Primitive neuroectoderm"
 /lab_host="DH10B (TI phase-resistant)"
 /note="Organ: brain; Vector: pNR-LIB (Clontech); Site_1
 SfiI (ggcgccctcgccg); Site_2: SfiI (ggccatcagcc);
 Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCCAATATGGCC-3' and 3' adaptor
 sequence: 5'-ATTGTAGGCGCGCGCCGACAAATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 215 c 458 g 225 t 4 others

376 a	215 c	458 g	225 t	4 others
-------	-------	-------	-------	----------

Query Match	69.0%;	Score 21.4;	DB 10;	Length 1278;
Best Local Similarity	80.6%;	Pred. No. 4e+03;		
Matches 25; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0

```

QY      1  ctgcgcgcgcctccgcgcctctgctgcg 31
          ||| ||| ||| ||| ||| ||| ||| |||
Db     1215  CTCGCGCGGCGCTCCCGCCCTCTCTCGGCG 1185

```

ECSE 28-NOV-2001

LOCUS	BG3633281	424 bp	mRNA	linear	EST 28-NOV-2001
DEFINITION	sac18f04.y1 Gm-cl051 glycine max cDNA clone				GENOME SYSTEMS CLONE
ID:	Gm-cl051-2576	5'	similar to	TR:Q9S778	Q9S778 T2E6.14. / mRNA sequence.

BG363281
BG363281.1
GI:13252378
ECF

soybean.
 glycine max
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
 Rosidae: eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae
 glycin.

1 (bases 1 to 424)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
Muller, T. Kueba M. Martin T. Beck C

A. A. Bolla, M. M. Maitra, M. M. Rattler, L. L. Novakova, J. J. Malling, C. C. Deane,
W. W. Underwood, K. K. Stepietoe, M. M. Theising, B. B. Allen, M. M. Bowers
Y. Y. Person, B. B. Swaller, T. T. Gibbons, M. M. Pape, D. D. Harvey, N. N. Schuck
R. R. Ritter, E. E. Kohn, S. S. Shin, T. T. Jackson, Y. Y. Cardenas, M. M. McCann
R. R. Waterston, R. R. and Wilson, R.

TITLE	JOURNAL	COMMENT
Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project

Location/Qualifiers
1. . 424

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4353 or contact via email: cc@resgen.com
High quality sequence, stop: 387.

100 a	106 c	94 g	124 t
-------	-------	------	-------

Query Match	67.7%	Score 21	DB 10	Length 424
Best Local Similarity	82.8%	Pred. No. 5.5e+03		
Matches 24	Conservative 0	Mismatches 5	Indels 0	Gaps 0

```

Oy      3  cgcgcggccctccgcgcctctgctcg 31
          ||| ||| ||| ||| ||| ||| |||
Db      13  cgcgcggcctcgccctcttcgacgacg 41

```

Search completed: August 14, 2002, 21:04:17
Job time: 11005 sec

Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
Location/Qualifiers
FEATURES 1..264
SOURCE /organism="Papio hamadryas"
/db_xref="taxon:9557"
gene <1..>264
/gene="SCA2"
/note="spino cerebellar ataxia 2"
BASE COUNT 25 a 130 c 78 g 31 t
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
Db 12 CCTCCCGCCCTTCGTCGTC 32

RESULT 2
AF330031 303 bp DNA linear PRI 08-NOV-2001
LOCUS AF330031 Macaca mulatta SCA2 gene, partial sequence.
DEFINITION AF330031
ACCESSION AF330031
VERSION AF330031.1 GI:12382833
KEYWORDS
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 303)
AUTHORS Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
DEFINITION 2 (bases 1 to 303)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
Location/Qualifiers
FEATURES 1..303
SOURCE /organism="Macaca mulatta"
/db_xref="taxon:9544"
gene <1..>303
/gene="SCA2"
/note="spino cerebellar ataxia 2"
BASE COUNT 32 a 143 c 92 g 36 t
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 303;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
Db 6 CCTCCCGCCCTTCGTCGTC 26

RESULT 3
AF330033 322 bp DNA linear PRI 08-NOV-2001
LOCUS AF330033 Macaca radiata SCA2 gene, partial sequence.
DEFINITION AF330033
ACCESSION AF330033
VERSION AF330033.1 GI:12382835

KEYWORDS bonnet macaque.
SOURCE Macaca radiata
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 322)
AUTHORS Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
DEFINITION 2 (bases 1 to 322)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
Location/Qualifiers
FEATURES 1..322
SOURCE /organism="Macaca radiata"
/db_xref="taxon:9548"
gene <1..>322
/gene="SCA2"
/note="spino cerebellar ataxia 2"
BASE COUNT 32 a 155 c 95 g 40 t
ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
Db 35 CCTCCCGCCCTTCGTCGTC 55

RESULT 4
ARI59544 355 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59544 Sequence 1 from patent US 6251589.
DEFINITION ARI59544
ACCESSION ARI59544
VERSION ARI59544.1 GI:16222225
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 355)
AUTHORS Tsuji,S. and Sempel,K.
TITLE Method for diagnosing spino cerebellar ataxia type 2 and primers
therefor
JOURNAL Patent: US 6251589-A 1 26-JUN-2001;
Location/Qualifiers
FEATURES 1..355
SOURCE /organism="unknown"
BASE COUNT 20 a 176 c 102 g 55 t 2 others
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
Db 158 CCTCCCGCCCTTCGTCGTC 178

RESULT 5
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS AF330030

```
DEFINITION   Presbytis entellus SCA2 gene, partial sequence.
ACCESSION    AF330030
VERSION      AF330030.1 GI:12382832
KEYWORDS
SOURCE       Hanuman langur.
ORGANISM     Presbytis entellus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
              Colobinae; Presbytis.
REFERENCE    1 (bases 1 to 384)
AUTHORS     Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
              Brahmachari,S.K.
TITLE        CAG repeat instability at SCA2 locus: anchoring CAA interruptions
              and linked single nucleotide polymorphisms
JOURNAL     Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED      11689490
REFERENCE    2 (bases 1 to 384)
AUTHORS     Choudhry,S. and Brahmachari,S.K.
TITLE        Direct Submission
JOURNAL     Submitted (21-DEC-2000) Functional Genomics Unit, Center for
              Biochemical Technology, Delhi University Campus, Mall Road, Delhi
              110 007, India
FEATURES     Location/Qualifiers
              1..384
              /organism="Presbytis entellus"
              /db_xref="taxon:9574"
              <1..>384
              /gene="SCA2"
              /note="spino cerebellar ataxia 2"
BASE COUNT   46 a 178 c 109 g 51 t
ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cctccgcccttcgtcgtc 21
Db 12 cctccgcccttcgtcgtc 32

RESULT 6
LOCUS       AF330028 390 bp DNA linear PRI 08-NOV-2001
DEFINITION Pan troglodytes SCA2 gene, partial sequence.
ACCESSION   AF330028
VERSION     AF330028.1 GI:12382830
KEYWORDS
SOURCE      chimpanzee.
ORGANISM    Pan troglodytes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE    1 (bases 1 to 390)
AUTHORS     Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
              Brahmachari,S.K.
TITLE        CAG repeat instability at SCA2 locus: anchoring CAA interruptions
              and linked single nucleotide polymorphisms
JOURNAL     Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED      11689490
REFERENCE    2 (bases 1 to 390)
AUTHORS     Choudhry,S. and Brahmachari,S.K.
TITLE        Direct Submission
JOURNAL     Submitted (21-DEC-2000) Functional Genomics Unit, Center for
              Biochemical Technology, Delhi University Campus, Mall Road, Delhi
              110 007, India
FEATURES     Location/Qualifiers
              1..390
              /organism="Pan troglodytes"
              /db_xref="taxon:9598"
              <1..>390
              /note="microsatellite"
              /rpt_type=tandem
repeat_region

DEFINITION   Presbytis entellus SCA2 gene, partial sequence.
ACCESSION    AF330030
VERSION      AF330030.1 GI:12382832
KEYWORDS
SOURCE       Hanuman langur.
ORGANISM     Presbytis entellus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
              Colobinae; Presbytis.
REFERENCE    1 (bases 1 to 384)
AUTHORS     Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
              Brahmachari,S.K.
TITLE        CAG repeat instability at SCA2 locus: anchoring CAA interruptions
              and linked single nucleotide polymorphisms
JOURNAL     Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED      11689490
REFERENCE    2 (bases 1 to 384)
AUTHORS     Choudhry,S. and Brahmachari,S.K.
TITLE        Direct Submission
JOURNAL     Submitted (21-DEC-2000) Functional Genomics Unit, Center for
              Biochemical Technology, Delhi University Campus, Mall Road, Delhi
              110 007, India
FEATURES     Location/Qualifiers
              1..384
              /organism="Presbytis entellus"
              /db_xref="taxon:9574"
              <1..>384
              /gene="SCA2"
              /note="spino cerebellar ataxia 2"
BASE COUNT   46 a 178 c 109 g 51 t
ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cctccgcccttcgtcgtc 21
Db 12 cctccgcccttcgtcgtc 32

RESULT 7
LOCUS       AF330029 409 bp DNA linear PRI 08-NOV-2001
DEFINITION Gorilla gorilla SCA2 gene, partial sequence.
ACCESSION   AF330029
VERSION     AF330029.1 GI:12382831
KEYWORDS
SOURCE      gorilla.
ORGANISM    Gorilla gorilla
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Gorilla.
REFERENCE    1 (bases 1 to 409)
AUTHORS     Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
              Brahmachari,S.K.
TITLE        CAG repeat instability at SCA2 locus: anchoring CAA interruptions
              and linked single nucleotide polymorphisms
JOURNAL     Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED      11689490
REFERENCE    2 (bases 1 to 409)
AUTHORS     Choudhry,S. and Brahmachari,S.K.
TITLE        Direct Submission
JOURNAL     Submitted (21-DEC-2000) Functional Genomics Unit, Center for
              Biochemical Technology, Delhi University Campus, Mall Road, Delhi
              110 007, India
FEATURES     Location/Qualifiers
              1..409
              /organism="Gorilla gorilla"
              /db_xref="taxon:9593"
              <1..>409
              /gene="SCA2"
              /note="spino cerebellar ataxia 2"
BASE COUNT   35 a 196 c 120 g 58 t
ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cctccgcccttcgtcgtc 21
Db 40 cctccgcccttcgtcgtc 60

RESULT 8
LOCUS       AR159558 572 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 18 from patent US 6251589.
ACCESSION   AR159558
VERSION     AR159558.1 GI:16222251
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE    Unclassified.
              1 (bases 1 to 572)
              Tsuji,S. and Sempel,K.
```

TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
 JOURNAL Patent: US 6251589-A 18-26-JUN-2001;
 FEATURES Location/Qualifiers
 source 1..572 /organism="unknown"
 BASE COUNT 34 a 277 c 174 g 85 t 2 others
 ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 572;
 Best Local Similarity 100.0%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0;

Oy 1 cctcccgcctcgtcgtc 21
 Db 158 CCTCCCCCGCCCTTCGTCGTC 178

RESULT 9
 LOCUS ARI59546 623 bp DNA linear PAT 17-OCT-2001
 DEFINITION Sequence 5 from patent US 6251589.
 ACCESSION ARI59546
 VERSION ARI59546.1 GI:16222229
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 623)
 AUTHORS Tsuji,S. and Sempel,K.
 TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
 JOURNAL Patent: US 6251589-A 5-26-JUN-2001;
 FEATURES Location/Qualifiers
 source 1..623 /organism="unknown"
 BASE COUNT 55 a 292 c 189 g 85 t 2 others
 ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 623;
 Best Local Similarity 100.0%; Pred. No. 99; Mismatches 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0;

Oy 1 cctcccgcctcgtcgtc 21
 Db 158 CCTCCCCCGCCCTTCGTCGTC 178

RESULT 10
 LOCUS HSDNSCA2 4163 bp mRNA linear PRI 09-JAN-1997
 DEFINITION H. sapiens mRNA for SCA2 protein.
 ACCESSION Y08262
 VERSION Y08262.1 GI:1770389
 KEYWORDS SCA2 gene.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 4163)
 AUTHORS Imbert,G., Saudou,F., Yvert,G., Devys,D., Trottier,Y., Garnier,J.M., Weber,C., Mandel,J.L., Cancel,G., Abbas,N., Duerr,A., Didierjean,O., Stevanin,G., Agid,Y. and Brice,A.
 TITLE Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/ glutamine repeats
 JOURNAL Nat. Genet. 14 (3), 285-291 (1996)
 MEDLINE 97051922
 REFERENCE 2 (bases 1 to 4163)
 AUTHORS Imbert,G.
 TITLE Direct Submission
 JOURNAL Submitted (20-SEP-1996) G. Imbert, I.G.B.M.C., Departement Of

FEATURES Genetics, B.P. 163, 67404 Illkirch Cedex, FRANCE
 source Location/Qualifiers
 1..4163
 /organism="Homo sapiens"
 /isolate="DAN patient"
 /db_xref="taxon:9606"
 /cell_line="lymphoblastoid"
 /clone_lib="DAN"
 /dev_stage="adult"
 1..2747
 /gene="SCA2"
 <1..2747
 /gene="SCA2"
 /codon_start=3
 /protein_id="CAA69589.1"
 /db_xref="GI:1770390"
 /db_xref="SPTREMBL:O99493"
 /translation="GNGGAFRPGSRRLGLGPPRPFFVLLPLASPGAPPAAPTRA
 SPGLARAPPSRGVSLARPAAPGCPRPACPEVYGLTWSLKPQQQQQQQQQQQQQQQ
 QQQQQPPPAANVRKPGSGILASPAAPSPSSSVSSSATAPSSVVAATSGGRPE
 LGRGRNSKGLPOSTISPDITAMRWVILTSYVSKCEVOYKNGIYGVEKTYSP
 KCDVLDAHKEKSTLESSSGCKRERIMESILFKCSDPYVNFKMDSSYAKRDAFTTSA
 ISAVNGEHKEKLEPMDAGELTANLELELVDSNGMDPNDMPRIENEGYVSTY
 DSSLSTYVPLERDNSEFLKREARAOLEIESSAQYARVALENDDESEKTTA
 VORNSEREHGSINTRENKTYIPGQRRREVISMGSRGNSPRMGQPGSGMPSTRSH
 TSDPNNGSDQRYVNGGVPMWSPCPSPSPSPSRYSYOGCPNSLPPRAAPTREPSP
 SRSPRPSP
 EFVSHNPPSP
 VLASFOAGITTEVAMPPIPAASPTPSPASPNRAVPTSSSKAKDRODONPACNK
 ENIRPNETSPSPSKAENKGISPVSEHKQIDLUKTKNPRLOPSPSTSSMOQLNK
 NREKGSRLDLKDIKIPSAKDSFIENSSSNTSGSSKPNSPSISPSLSTENKRGPE
 VTSQGVOTSPSPACQEKDKDEKDAEYVKRSTYLNPNNAEFNPRFSQKPTTPPS
 PPQAQPSPMVGHQPPVYTGQVCFAPNMYPVVSPGVQYOICPNSGKTSIIRVP"
 BASE COUNT 1136 a 1196 c 908 g 923 t
 ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 4163;
 Best Local Similarity 100.0%; Pred. No. 64; Mismatches 0; Indels 0; Gaps 0;
 Matches 21; Conservative 0;

Oy 1 cctcccgcctcgtcgtc 21
 Db 60 CCTCCCCCGCCCTTCGTCGTC 80

RESULT 11
 LOCUS A62706 4200 bp DNA linear PAT 12-MAR-1998
 DEFINITION Sequence 7 from patent WO9717445.
 ACCESSION A62706
 VERSION A62706.1 GI:3716590
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 4200)
 AUTHORS Toral,L., Lutz,Y., Trottier,Y., Mandel and Jean-Louis.
 TITLE METHOD FOR TREATING NEURODEGENERATIVE DISEASES USING A 102 ANTIBODY OR A FRAGMENT OR DERIVATIVE THEREOF, AND CORRESPONDING PHARMACEUTICAL COMPOSITIONS
 JOURNAL Patent: WO 97/17445-A 7 15-MAY-1997;
 COMMENT CENTRE NAT RECH SCIENT (FR)
 FEATURES Other publication FR 2741088 19970516.
 source Location/Qualifiers
 1..4200
 /organism="unidentified"
 /db_xref="taxon:32644"
 /clone="DAN1"
 BASE COUNT 1152 a 1200 c 913 g 935 t
 ORIGIN

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 231758)
Morley, K.C.
Direct Submission
Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced gi:966929.

Genome Center
Center: Baylor College of Medicine
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: UG
Center clone name: RP11-42B1

Assembly program: Phrap; version 0.990329
Consensus quality: 224788 bases at least Q40
Consensus quality: 229074 bases at least Q30
Consensus quality: 230948 bases at least Q20
Estimated insert size: 227317; sum-of-contigs estimation
Estimated insert size: 317311; agarose-fp estimation
Quality coverage: 6.3x in Q20 bases; agarose-fp estimation
Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 20 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of 'N', but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 33241: contig of 33241 bp in length
* 33242 33341: gap of unknown length
* 33342 56391: contig of 23050 bp in length
* 56392 56491: gap of unknown length
* 56492 81323: contig of 24832 bp in length
* 81324 81423: gap of unknown length
* 81424 102538: contig of 21115 bp in length
* 102539 102639: gap of unknown length
* 102639 119710: contig of 17072 bp in length
* 119711 119810: gap of unknown length
* 119811 136913: contig of 17103 bp in length
* 136914 137013: gap of unknown length
* 137014 153285: contig of 16272 bp in length
* 153286 153385: gap of unknown length
* 153386 167987: contig of 14602 bp in length

167988 168087: gap of unknown length
* 168088 178731: contig of 10644 bp in length
* 178732 178731: gap of unknown length
* 178732 186641: contig of 7810 bp in length
* 186642 186741: gap of unknown length
* 186741 193215: contig of 6474 bp in length
* 193216 193315: gap of unknown length
* 193316 201310: contig of 7993 bp in length
* 201311 201410: gap of unknown length
* 201411 208647: contig of 7237 bp in length
* 208648 208747: gap of unknown length
* 208748 213802: contig of 5055 bp in length
* 213803 213902: gap of unknown length
* 213903 218049: contig of 4147 bp in length
* 218050 218149: gap of unknown length
* 218150 223416: contig of 5167 bp in length
* 223417 227389: contig of 3973 bp in length
* 227390 227489: gap of unknown length
* 227490 229032: contig of 1543 bp in length
* 229033 229147: gap of unknown length
* 229148 230651: contig of 1519 bp in length
* 230652 230751: gap of unknown length
* 230752 231758: contig of 1007 bp in length.

Location/Qualifiers
1. 231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

Query Match 100.0%; Score 21; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cctcccccgcctcgtcgc 21
|||||
Db 89326 CCTCCCCGCCCTTCGTCGTC 89306

RESULT 15
LOCUS AF041472 4225 bp mRNA linear ROD 28-NOV-2001
DEFINITION Mus musculus ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION AF041472
VERSION AF041472.1 GI:3005019

KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4225)
AUTHORS Nechiporuk, T.T., Huyuh, D.P., Figueroa, K., Sahba, S., Nechiporuk, A.V.
and Pulst, S.M.

TITLE The mouse SCA2 gene: cDNA sequence, alternative splicing and
protein expression
JOURNAL Hum. Mol. Genet. 7 (8), 1301-1309 (1998)

MEDLINE 98334550
PUBMED 9668173
REFERENCE 2 (bases 1 to 4225)
AUTHORS Nechiporuk, T.T., Figueroa, K., Sahba, S., Nechiporuk, A.V. and
Pulst, S.M.

TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Medicine/Neurology, Cedars-Sinai Medical
Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
Location/Qualifiers
1. 4225
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="12q23.1"

gene
CDS

1..4225
/gene="SCA2"
27..3884
/gene="SCA2"
/codon_start=1
/product="ataxin-2"
/protein_id="AAC09275.1"
/db_xref="GI:3005020"

/translation="MRSSTAAYORPAAGDPEPRRPAAGNARSLPRTARRGGGGA
VPSAGPPPRGAPPRGRSPPCASDCRSGNGHGASRGRSLRGCGPRPFVYLL
ALAPATPARACPGVRASPPRSGVSSSAPAPGCPRACEPVYGLTMSLKPPQPP
APATGRKPGGGLSSPGAAPASAAVTSASVVPAPAPVASSAAGGRRGLGRNS
SKGLPPTISFDGIYANVMVHILTSVSGKCEVOKNGIYEGVFKTYSKCDLVD
AAHEKSTESSGPKREEIMESVLFKCSDEVVYOFKDTSSYARDAFTDSALAKNG
EHKEDLPMDAGELTASFELELNDVSNQMDPMERNENGVSTYDSSLSTY
VPLERDNSEETLKREARANQLAEIETESSNQYKARVALENDDSBEKTYAVRNCSDR
EGHGPTRDNKTYIPQQRNREVLWSGKQSSPRMGQPGGMPSPRAASHISDFNPA
GSDORVYNGVPMSPCPSHSRPPSRYSQSPNSLPPRAATHTRPSPRSPSRPS
HPSAHGSPAPVSTMPKRMSESGPPRMSPKQRRHPRNHRYAGSGMSGLFEVSHNP
SEAAPPVARTSPAGTMSVSGVRLSPKTHRPSPROSSIGNSPSGVLASPOAG
ITPAEAVSMVPVPAASPTPASPSNRALTPTSEAKDSRLQDORONSPAGSKENYKSET
SPFSKADNKGKSPVYSEHRKQIDDLKFKKNDPRLOPSTSPMDQLSKNREGESR
DLIKDKTENSAKDSFTIDSSSSSNTSGSSKTNSPSTSPSLSNAEHKGREYTSQGV
QTSSPACQKQEKDREKKDTQVRKSTLNPNAKENPRSFQPKFTTPTSPRQQAQ
PSPSMVGHQOPAVYTQVCFAPNNMYPVVSPGVOLYPIPTMPMPVNOAKTYRAGK
VNNMPOQRDOHQSMTMHHPASAGPPIVATPPAYSTQYVAYSPQFPNQPLVOHVPH
YOSHPHYVSPVIOGNARMAPPAHQAQGLVSSAAQFGAHEOTHAMYACPKLPYKE
TSPSYFAISTGSLAQOYVAPNAALHPHTPHOPASATPTGGOOSOHGSGHPAPSPYOH
HGHQAQAALHLASPOQOSHIYHAGLAPTTPSMTPASNTOSPOSSFPAAQOYVFTIIRS
HVQPAATTPPHMAHVQAHVQSGVSPSHPTAHAPMMLMTTOPPGKAAALQSHLOPITP
VSTTAHFPYMTHPSVQAHHQOL"

BASE COUNT 1007 a 1324 c 1042 g 851 t 1 others
ORIGIN

Query Match 95.2%; Score 20; DB 10; Length 4225;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctcccgcccttgctgcgt 20
|||||
Db 303 cctcccgcccttgctgcgt 322

Search completed: August 14, 2002, 21:48:25
Job time: 13523 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 22:06:34 ; Search time 906.46 Seconds
(Without alignments)
39.776 Million cell updates/sec

Title: US-09-707-919-5
Perfect score: 1 cctcccgcctcctgcctc 21
Sequence: 1 cctcccgcctcctgcctc 21

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	21	100.0	355	19	AAV17224
2	21	100.0	516	19	AAV06551
3	21	100.0	623	19	AAV17229
4	21	100.0	4200	18	AAV78912
5	21	100.0	4367	19	AAV30270
6	21	100.0	4481	19	AAV06552
7	21	100.0	4481	20	AAZ23428
8	17.8	84.8	6042	24	ABL33593
9	16.4	78.1	277	14	AAO39964

10	16.4	78.1	1632	14	AAO39968
11	16.4	78.1	1632	14	AAO39969
12	16.2	77.1	115	22	AAK69214
13	16.2	77.1	250	22	AAK69215
14	16.2	77.1	361	22	AAK69216
15	16.2	77.1	377	21	AAK69217
16	16.2	77.1	402	21	AAK69218
17	16.2	77.1	435	21	AAK69219
18	16.2	77.1	468	21	AAK69220
19	16.2	77.1	581	22	AAK69221
20	16.2	77.1	588	22	AAK69222
21	16.2	77.1	711	22	AAK69223
22	16.2	77.1	784	20	AAV69491
23	16.2	77.1	1627	22	AAK69217
24	16.2	77.1	1743	18	AAV69491
25	16.2	77.1	1743	21	AAV69491
26	16.2	77.1	2316	14	AAO50419
27	16.2	77.1	5270	24	ABL32734
28	16.2	77.1	5270	24	ABL32734
29	16.2	77.1	7361	19	AAV62153
30	16.2	77.1	9775	20	AAV77721
31	16.2	77.1	9934	20	AAV77722
32	16.2	77.1	15698	24	ABL34141
33	16.2	77.1	20510	23	ABL02872
34	16.2	77.1	21010	22	ABL05888
35	16.2	77.1	21010	22	AAK89247
36	16.2	77.1	21024	22	AAK89248
37	16.2	77.1	21024	22	AAK89249
38	16.2	77.1	130	21	AAV10847
39	16.2	77.1	8246	23	AAV59621
40	15.8	75.2	384	23	AAV6465
41	15.8	75.2	491	23	AAV68077
42	15.8	75.2	2704	23	ABL13675
43	15.8	75.2	8571	22	AAV19406
44	15.8	75.2	8571	22	AAV19406
45	15.8	75.2	11138	22	AAV19405

ALIGNMENTS

RESULT 1	
AAV17224	
ID AAV17224 standard; DNA: 355 BP.	
XX AAV17224;	
AC 29-JUN-1998 (first entry)	
XX 29-JUN-1998	
XX SCA2 gene fragment.	
XX SCA2 gene fragment.	
XX SCA2 gene: spinocerebellar ataxia type II; CAG repeat; PCR primer; ss.	
XX Synthetic.	
XX Key	Location/Qualifiers
XX CDS	341..355
XX FT	/*tag= a
XX FT	/note= "SCA2 protein fragment"
XX PN	W09803679-A1.
XX PD	29-JAN-1998.
XX PF	18-JUL-1996; 96WO-JP01999.
XX PR	18-JUL-1996; 96WO-JP01999.
XX PA	(SRLS-) SRL INC.
XX PI	Sanpel K, Tsuji S;
XX DR	WPI: 1998-120796/11.

PKC-gamma DNA. Ra
PKC-gamma promoter
Human immune/haema
Human immune/haema
Novel human polynu
Plant microsatelli
Human immune/haema
Plant microsatelli
Plant microsatelli
Human immune/haema
Human nervous syst
Chinese hamster fu
Banana fruit ripen
Human immune/haema
Human immune/haema
Human TULP4 CDNA.
Mouse TULP4 CDNA.
Partial sequence o
Human immune syste
Human gene regulat
HSV-2 strain SB5 C
N. crassa his-3 co
N. crassa his-3 co
Human immune syste
Drosophila melanog
Human reproductive sy
Human digestive sy
Human digestive sy
Fusarium venenatum
Propionibacterium
Novel human polynu
DNA encoding novel
Drosophila melanog
Human excretory re
Human kidney relat
Human excretory re

DR	P-PSDB; AAM41370.
XX	Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT	number of CAG repeat units
XX	
PS	Claim 1; Page 10; 23pp; Japanese.
XX	
CC	This sequence represents a fragment of the SCA2 gene. It can be used in
CC	the method of the invention for diagnosing spinocerebellar ataxis type
CC	II, by performing PCR on the test DNA using two primers hybridising to
CC	parts of the SCA2 gene sequence, and determining the number of CAG
CC	repeats in the amplified products. The method provides an easy means for
CC	the diagnosis of spinocerebellar ataxis type II.
XX	
SQ	Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;
Oy	Query Match 100.0%; Score 21; DB 19; Length 355; Best Local Similarity 100.0%; Pred. No. 7; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Db	1 cctcccgcccttcgtcgc 21 158 cctcccgcccttcgtcgc 178
RESULT 2	
AAV06551	AAV06551 standard; DNA; 516 BP.
XX	
AC	AAV06551;
XX	
DT	06-JUL-1998 (first entry)
XX	
DE	SCA2 gene fragment including CAG repeat region.
KM	SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
KW	diagnosis; olivoponto-cerebellar atrophy; ss; ds.
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	primer_bind complement (241..257) /*tag= d
FT	primer_bind /note= "primer SCA2-A binding site"
FT	primer_bind 349..366 /*tag= b
FT	exon /note= "primer SCA2-B binding site"
FT	exon 499..500 /*tag= C
FT	repeat_region /note= "predicted splice site"
FT	repeat_region 267..332 /*tag= d
FT	repeat_unit /note= "CAG repeat region"
FT	repeat_unit 267..269 /*tag= e
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 270..272 /*tag= f
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 273..275 /*tag= g
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 276..278 /*tag= h
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 279..281 /*tag= i
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 282..284 /*tag= j
FT	repeat_unit /note= "CAG repeat"
FT	repeat_unit 285..287

FT	/*tag=	k	
FT	/note=	"CAG repeat"	
FT	repeat_unit	291..293	
FT	/*tag=	l	
FT	/note=	"CAG repeat"	
FT	repeat_unit	294..296	
FT	/*tag=	m	
FT	/note=	"CAG repeat"	
FT	repeat_unit	297..299	
FT	/*tag=	n	
FT	/note=	"CAG repeat"	
FT	repeat_unit	300..302	
FT	/*tag=	o	
FT	/note=	"CAG repeat"	
FT	repeat_unit	306..308	
FT	/*tag=	p	
FT	/note=	"CAG repeat"	
FT	repeat_unit	309..311	
FT	/*tag=	q	
FT	/note=	"CAG repeat"	
FT	repeat_unit	312..314	
FT	/*tag=	r	
FT	/note=	"CAG repeat"	
FT	repeat_unit	313..317	
FT	/*tag=	s	
FT	/note=	"CAG repeat"	
FT	repeat_unit	318..320	
FT	/*tag=	t	
FT	/note=	"CAG repeat"	
FT	repeat_unit	321..323	
FT	/*tag=	u	
FT	/note=	"CAG repeat"	
FT	repeat_unit	324..326	
FT	/*tag=	v	
FT	/note=	"CAG repeat"	
FT	repeat_unit	327..329	
FT	/*tag=	w	
FT	/note=	"CAG repeat"	
FT	repeat_unit	330..332	
FT	/*tag=	x	
FT	/note=	"CAG repeat"	
XX	W09742314-A1.		
PN			
PD	13-NOV-1997.		
XX			
XX	08-MAY-1997;	97WO-US07725.	
XX			
PR	08-OCT-1996;	96US-0727084.	
PR	08-MAY-1996;	96US-0017388.	
PR	19-JUL-1996;	96US-0022207.	
XX			
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.		
XX			
XX	Pulst S;		
XX			
DR	WPI; 1998-086523/08.		
XX			
PT	Nucleic acids encoding human and mouse ataxin 2 - a product of the		
PT	spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of the		
PT	ataxia type 2		
XX			
PS	Example 2; Page 51-52; 98pp; English.		
XX			
CC	This genomic DNA in plasmid pL65122B includes a CAG repeat region		
CC	from the novel human SCA2 gene (see AAV06552). It was identified		
CC	following the construction of a bacterial artificial chromosome		
CC	contig and a p1 artificial chromosome of the spinocerebellar		
CC	ataxia 2 (SCA2) gene region and the identification of the SCA2		
CC	gene from this contiguous map unit using a technique that screens		
CC	for the presence of DNA trinucleotide repeats. The SCA2 locus is		
CC	at 12q24.1. Ataxia type 2 can be diagnosed by detecting a genomic		
CC	or transcribed mRNA sequence in an individual having an expanded		

CC CAG repeat at a location corresponding to the CAG repeat region of
 CC the SCA2 gene. The presence of at least 13 CAG repeats above the
 CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
 CC primers (see AAT99640-41) amplifying at least this region are used
 CC for diagnosis. Also claimed are full-length ataxin-2 cDNAs for
 CC human and mouse (see AAV06552-53). Kits for detecting mutations at
 CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
 CC useful for studying the physiological roles of SCA2 polypeptide
 CC (ataxin-2, see AAM33807-08) and its effect upon behaviour.
 CC
 XX
 SO Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;

Query Match 100.0%; Score 21; DB 19; Length 516;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctccgccctcgtcgtc 21
 ||||||||||||||||
 Db 69 cctccgccctcgtcgtc 89

RESULT 3

AAV17229
 ID AAV17229 standard; DNA; 623 BP.

AC AAV17229;

DT 29-JUN-1998 (first entry)

DE SCA2 gene fragment.

KW SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

OS Synthetic.

FH Key Location/Qualifiers

FT CDS 341..583

FT /tag= a /note= "SCA2 protein fragment, no stop codon given"

FT W09803679-A1.

PD 29-JAN-1998.

PF 18-JUL-1996; 96WO-JP01999.

PR 18-JUL-1996; 96WO-JP01999.

PA (SRLS-) SRL INC.

PI Sanpei K, Tsuji S;

DR WPT: 1998-120796/11.

DR P-PSDB: AAM41372.

PT Diagnosing spinocerebellar ataxis type II - by PCR and determining
 number of CAG repeat units

PS Example 1; Page 11-12; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in
 CC the method of the invention for diagnosing spinocerebellar ataxis type
 CC II, by performing PCR on the test DNA using two primers hybridising to
 CC parts of the SCA2 gene sequence, and determining the number of CAG
 CC repeats in the amplified products. The method provides an easy means for
 CC the diagnosis of spinocerebellar ataxis type II.

SO Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 100.0%; Score 21; DB 19; Length 623;
 Best Local Similarity 100.0%; Pred. No. 6.8;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 cctccgccctcgtcgtc 21
 ||||||||||||||||
 Db 158 cctccgccctcgtcgtc 178

RESULT 4

AAT78912
 ID AAT78912 standard; cDNA; 4200 BP.

AC AAT78912;

DT 09-FEB-1998 (first entry)

DE Spinocerebellar ataxia gene SCA2.

KW Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;
 KW repeat region; affinity; TATA binding protein; Kennedy disease;
 KW transcription initiation factor; lymphoblastic cell line; schizophrenia;
 KW Huntington's disease; dominant autosomal spinocerebellar ataxia;
 KW X-linked spinocerebellar muscular atrophy; familial spastic paraplegia;
 KW dentatorubral-pallidolusial atrophy; bipolar affective disorder;
 KW manic depressive psychosis; ss.

OS Homo sapiens.

FH Key

FT CDS 3..2747

FT /tag= a /note= "this CDS contains a putative translational start
 codon for the SCA2 protein at positions 243-245"

FT CDS 2594..3640

FT /tag= b /note= "this second open reading frame may be derived
 by a frameshift or by alternative splicing"

FT CDS 3..242

FT /tag= c /note= "putative open reading frame which is in frame
 with the putative translational start site of
 the SCA2 open reading frame"

FT CDS 239..245

FT /tag= d /note= "putative Kozak consensus signal"

FT repeat_region 258..323

FT /tag= e /note= "encodes polyglutamine repeat region; contains
 repeats of CAG with 2 CAA codons interspersed"

FT repeat_unit 258..260

FT /tag= f /note= "CAG repeats"

FT misc_feature 1..3986

FT /tag= g /note= "sequence contained in DAN1 clone"

FT misc_feature 3987..4200

FT /tag= h /note= "derived from the EST's AAH92640, AAN90240 and
 AAZ13574 from dbEST database"

FT misc_feature 4023..4029

FT /tag= i /note= "region which differs in length between the
 sequences of the EST clones AAH92640, AAN90240
 and AAZ13574"

FT W09717445-A1.

PD 15-MAY-1997.

PF 08-NOV-1996; 96WO-FR01773.

PR 10-NOV-1995; 95FR-0013576.

PA (CNRS) CNRS CENT NAT RECH SCI.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Lutz Y, Mandel J, Tora L, Trottier Y;
XX
DR WPI: 1997-281034/25.
DR P-PSDB: AAW24800, AAW24801.
XX
PT Antibody 1C2 used for treating or preventing neuro-degenerative
PT diseases - associated with proteins containing long poly:glutamine
PT repeats, e.g. Huntington's disease
XX
PS Claim 21: Page 45-47: 69pp: French.
XX
CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
CC treatment of neurodegenerative diseases associated with the presence
CC of polyglutamine repeat regions. This Mab is already known for its
CC affinity to the TATA binding protein (TBP) transcription initiation
CC factor, especially at the amino acid sequence LEEQORQ00000 found at
CC the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
CC for polyglutamine repeats with a proportional affinity to the number
CC of glutamine repeats. This affinity has been used to identify genes
CC encoding proteins containing long polyglutamine repeats which are
CC implicated in neurodegenerative diseases. A screen of an expression
CC library, generated from a lymphoblastic cell line from a patient
CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
CC new sequences (AA78906-T78911) encoding polyglutamine repeats. Mab 1C2
CC also isolated the complete SCA2 gene in clone DAN1 (sequence presented
CC here). The sequence appears to contain 2 open reading frames (ORF) the
CC second of which may be generated by an frameshift slippage or by an
CC alternative splicing event. The first ORF also encodes a 22 amino acid
CC polyglutamine repeat region near the N-terminus of the protein. Normal
CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CMA repeats
CC interspersed whereas the mutant sequence from patients with SCA
CC contains at least 30, preferably 37-50 CAG repeats.
CC Mab 1C2, active fragment of it or nucleic acids encoding it are
CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
CC X-linked spinobulbar muscular atrophy (Kennedy disease),
CC denatorubral-pallidolusial atrophy, dominant autosomal spinocerebellar
CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
CC depressive psychoses and schizophrenia.
XX
SQ Sequence 4200 BP: 1152 A: 1200 C: 913 G: 935 T: 0 other;

Query Match 100.0%; Score 21; DB 18; Length 4200;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
DB 60 cctcccgcccttgctgc 80

RESULT 5
AAV30270
ID AAV30270 standard; DNA: 4367 BP.
XX
AC AAV30270;
XX
DT 02-OCT-1998 (first entry)
XX
DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
XX
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 49..3990
FT /*tag= a
FT /product= "Spinocerebellar ataxia type 2 associated

FT repeat_region 544..612 ProteinN
FT /*tag= b
FT /note= "normal CAG repeat region; this is increased in
FT repeat_unit 544..546 Patients with SCA2"
FT /*tag= c
XX
PN W09818920-A1.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-JP03946.
XX
PR 30-OCT-1996; 96UP-0304059.
XX
PA (SRLS-) SRL INC.
XX
PI Sanpei K, Tsuji S;
XX
DR WPI: 1998-272215/24.
DR P-PSDB: AAW60213.
XX
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
PS Claim 1: Pages 13-22: 38pp: Japanese.
XX
CC This represents the sequence of a gene causative of spinocerebellar
CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
XX
SQ Sequence 4367 BP: 1124 A: 1328 C: 991 G: 924 T: 0 other;

Query Match 100.0%; Score 21; DB 19; Length 4367;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgcccttgctgc 21
|||||
DB 346 cctcccgcccttgctgc 366

RESULT 6
AAV06552
ID AAV06552 standard; cDNA: 4481 BP.
XX
AC AAV06552;
XX
DT 06-JUL-1998 (first entry)
XX
DE Human SCA2 cDNA including CAG repeat region.
XX
KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 164..4101
FT /*tag= a
FT primer_bind complement (631..648)
FT /*tag= b

```

FT /note= "primer SCA2-A binding site"
FT primer_bind 740..757
FT /tag= C
FT /note= "primer SCA2-B binding site"
FT primer_bind 1070..1091
FT /tag= d
FT /note= "primer SCA2-14B binding site"
FT exon 899..900
FT /tag= e
FT /note= "predicted splice site"
FT 658..723
FT /tag= f
FT /note= "CAG repeat region"
FT repeat_unit 658..660
FT /tag= g
FT /note= "CAG repeat"
FT 661..663
FT /tag= h
FT /note= "CAG repeat"
FT repeat_unit 664..666
FT /tag= i
FT /note= "CAG repeat"
FT 667..669
FT /tag= j
FT /note= "CAG repeat"
FT repeat_unit 670..672
FT /tag= k
FT /note= "CAG repeat"
FT 673..675
FT /tag= l
FT /note= "CAG repeat"
FT repeat_unit 676..678
FT /tag= m
FT /note= "CAG repeat"
FT repeat_unit 679..681
FT /tag= n
FT /note= "CAG repeat"
FT 685..687
FT /tag= o
FT /note= "CAG repeat"
FT repeat_unit 688..690
FT /tag= p
FT /note= "CAG repeat"
FT repeat_unit 691..693
FT /tag= q
FT /note= "CAG repeat"
FT repeat_unit 694..696
FT /tag= r
FT /note= "CAG repeat"
FT 700..702
FT /tag= s
FT /note= "CAG repeat"
FT repeat_unit 703..705
FT /tag= t
FT /note= "CAG repeat"
FT 706..708
FT /tag= u
FT /note= "CAG repeat"
FT repeat_unit 709..711
FT /tag= v
FT /note= "CAG repeat"
FT 712..714
FT /tag= w
FT /note= "CAG repeat"
FT repeat_unit 715..717
FT /tag= x
FT /note= "CAG repeat"
FT repeat_unit 718..720
FT /tag= y
FT /note= "CAG repeat"
FT repeat_unit 721..723
FT /tag= z
FT /note= "CAG repeat"

```

```

XX MO9742314-A1.
XX
XX 13-NOV-1997.
XX
XX 08-MAY-1997; 97WO-US07725.
XX
XX 08-OCT-1996; 96US-0727084.
XX 08-MAY-1996; 96US-0017388.
XX 19-JUL-1996; 96US-0022207.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
XX Pulst S:
XX
XX WPI: 1998-086523/08.
XX
XX P-PSDB; AAW33807.
XX
XX Nucleic acids encoding human and mouse ataxin 2 - a product of the
XX spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
XX ataxia type 2
XX
XX
XX Claim 6; Page 52-58; 98pp; English.
XX
XX This cDNA sequence corresponds to a novel SCA2 gene encoding a human
XX spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
XX (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
XX human frontal cortex cDNA library in lambda ZapII were screened
XX with probes obtained by PCR amplification of plasmid AAP651228 (see
XX AAW05551). PCR products were used to screen the human adult frontal
XX cortex library and 5' clones were obtained by RT-PCR of placental
XX mRNAs. Overlapping clones were used to generate the composite 4481
XX bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
XX or transcribed mRNA sequence in an individual having an expanded
XX CAG repeat at a location corresponding to the CAG repeat region of
XX the SCA2 gene. The presence of at least 13 CAG repeats above the
XX normal level (22, occasionally 23, repeats) is indicative of SCA2.
XX Primers (see AAT99640-41) amplifying at least this region are used
XX for diagnosis. Also claimed are kits for detecting mutations at
XX the SCA2 locus, antisense oligonucleotides, and transgenic animals
XX useful for studying the physiological roles of ataxin-2 and its
XX effect upon behaviour.
XX
XX Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other:
XX
XX
XX Query Match 100.0%; Score 21; DB 19; Length 4481;
XX Best Local Similarity 100.0%; Pred. No. 6;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 cctcccgcccttcgctgc 21
XX ||||||||||||||||||||
XX Db 460 cctcccgcccttcgctgc 480
XX
XX RESULT 7
XX AA223428
XX ID AA223428 standard; DNA: 4481 BP.
XX
XX AC AA223428;
XX
XX DT 19-JAN-2000 (first entry)
XX
XX DE Human SCA2 DNA.
XX
XX KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
XX huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
XX atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
XX Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
XX dentatorubropallidolysian atrophy; cell proliferation; cell survival;
XX neoplastic; malignant; autoimmune; fibrotic; ss.
XX
XX OS Homo sapiens.

```

```

XX  Key      Location/Qualifiers
FT  CDS      163..4101
FT          /*tag= a
FT          /product= "SCA2"
XX
XX  MO9945944-A1.
XX
XX  16-SEP-1999.
XX
XX  11-MAR-1999; 99MO-US05250.
XX
XX  12-MAR-1998; 98US-0041886.
XX
XX  (BURN-) BURNHAM INST.
XX
XX  Bredesen DE, Rabizadeh S;
XX
XX  WPI; 1999-561617/47.
XX  P-PSDB; AAY33495.
XX
XX  New proapoptotic dependence peptides, used to develop products for
XX  treating, e.g. Alzheimer's disease -
XX
XX  Disclosure; Page 130-135; 1999p; English.
XX
XX  This invention describes novel pure proapoptotic dependence peptides
XX  which comprise a sequence of an active dependence domain selected from
XX  dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
XX  huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
XX  SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
XX  of inducing cell death and can be used to develop products to mediate or
XX  inhibit apoptosis. The methods can be used for reducing the severity of
XX  a proapoptotic dependence domain mediated pathological conditions e.g.
XX  Huntington's disease, Alzheimer's disease, Kennedy's disease,
XX  Spino cerebellar ataxias, dentatorubropallidoluysian atrophy,
XX  Machado-Joseph disease, stroke or head trauma. They can also be used for
XX  reducing the severity of a pathological condition mediated by upregulated
XX  cell proliferation or cell survival e.g. neoplastic, malignant,
XX  autoimmune or fibrotic conditions. This sequence encodes the human
XX  SCA2 polypeptide described in the method of the invention.
XX
XX  Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;
XX
XX  Query Match      100.0%; Score 21; DB 20; Length 4481;
XX  Best Local Similarity 100.0%; Pred. No. 6;
XX  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  0Y      1 cctcccccgccttcgtcgc 21
XX          |||||
XX  Db      460 cctcccccgccttcgtcgc 480
XX
XX  RESULT 8
XX  ABL33593/C
XX  ID      ABL33593 standard; DNA; 6042 BP.
XX
XX  AC      ABL33593;
XX
XX  DT      26-MAR-2002 (first entry)
XX
XX  DE      Human immune system associated gene SEQ ID NO: 1566.
XX
XX  KW      Human; immune system disease; cytosine methylation; antiasthmatic;
XX  antiarteriosclerotic; antihaemic; cytosratic; nocropic;
XX  neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX  antineumatic; antiarthritic; antidiabetic; antipsoriatic;
XX  antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX  acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX  neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX  gene; ds.

```

```

OS  Homo sapiens.
XX
XX  MO200200928-A2.
XX
XX  03-JAN-2002.
XX
XX  02-JUL-2001; 2001MO-EP07537.
XX
XX  30-JUN-2000; 2000DE-1032529.
XX  01-SEP-2000; 2000DE-1043826.
XX
XX  (EPIC-) EPIGENOMICS AG.
XX
XX  Olek A, Piepenbrock C, Berlin K;
XX
XX  WPI; 2002-130909/17.
XX
XX  Nucleic acid comprising fragment of chemically modified gene, useful
XX  for diagnosis and treatment of diseases associated with abnormal
XX  cytosine methylation -
XX
XX  Claim 1; SEQ ID NO 1566; 32pp + Sequence Listing; German.
XX
XX  The present invention provides a number of human immune system associated
XX  genes which are modified by the methylation of cytosines. The sequences
XX  can be used in the diagnosis and treatment of immune system disorders,
XX  including eye diseases such as retinopathy, neovascular glaucoma and
XX  macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX  leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX  rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX  diseases. The present sequence is a gene of the invention.
XX
XX  Sequence 6042 BP; 1371 A; 254 C; 1635 G; 2782 T; 0 other;
XX
XX  Query Match      84.8%; Score 17.8; DB 24; Length 6042;
XX  Best Local Similarity 90.5%; Pred. No. 1.2e+02;
XX  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX  0Y      1 cctcccccgccttcgtcgc 21
XX          |||||
XX  Db      617 CcTCCCGCCGCCCTCGTCGC 597
XX
XX  RESULT 9
XX  AAQ39964
XX  ID      AAQ39964 standard; DNA; 277 BP.
XX
XX  AC      AAQ39964;
XX
XX  DT      28-JUL-1993 (first entry)
XX
XX  DE      PKC-gamma promoter.
XX
XX  KW      Promoter; rat; protein kinase C; gamma; alpha; PKC; recombinant;
XX  vector; cranial; nerve cell; gene function; research; therapy;
XX  brain disease; ss.
XX
XX  OS      Rattus rattus.
XX
XX  PN      JP05056781-A.
XX
XX  PD      09-MAR-1993.
XX
XX  PF      25-DEC-1990; 90JP-0405849.
XX
XX  PR      26-DEC-1989; 89JP-0334751.
XX
XX  PA      (TAKE ) TAKEDA CHEM IND LTD.
XX
XX  DR      WPI; 1993-120380/15.
XX
XX  New expression vector contg. protein kinase C promoter - useful

```

PT	for expressing genes in cranial nerve cell for treating related
PT	disease
XX	
PS	Claim 5; Page 34; 39pp; Japanese.
XX	
CC	The sequences given in AA039964-65 represent promoter regions isolated
CC	from the rat protein kinase C (PKC) gamma and alpha genes respectively.
CC	These promoter sequences can be operatively linked to a structural
CC	gene in the production of a recombinant vector. This vector may be
CC	used in the production of transformants. Recombinant genes bearing
CC	these promoters can be used to express many genes in cranial nerve
CC	cells. This is useful in the research of gene function and for
XX	research and therapy of brain disease.
XX	
SQ	Sequence 277 BP; 54 A; 74 C; 98 G; 51 T; 0 other;
QY	4 cccgcacctgtcgtc 21
Db	175 cccgcacctgtcgtc 192
RESUT 10	
AA039968	
ID	AA039968 standard; DNA; 1632 BP.
XX	
AC	AA039968;
XX	
DT	28-JUL-1993 (first entry)
XX	
DE	PKC-gamma DNA.
XX	
KW	Promoter; rat; protein kinase C; gamma; alpha; PKC; recombinant;
KW	vector; cranial; nerve cell; gene function; research; therapy;
KW	brain disease; upstream region; ss.
XX	
OS	Rattus rattus.
XX	
FH	Key Location/Qualifiers
FT	Promoter 1356..1632
FT	/*tag= a
FT	/note= "PKC-gamma promoter"
XX	
PN	JP05056781-A.
XX	
PD	09-MAR-1993.
XX	
PF	25-DEC-1990; 90JP-0405849.
XX	
PR	26-DEC-1989; 89JP-0334751.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
DR	WPI; 1993-120380/15.
XX	
PT	New expression vector contg. protein kinase C promoter - useful
PT	for expressing genes in cranial nerve cell for treating related
PT	disease
XX	
PS	Disclosure; Page 37; 39pp; Japanese.
XX	
CC	The sequences given in AA039968-69 represent the promoter and associated
CC	upstream regions isolated from the rat protein kinase C (PKC) gamma
CC	gene. The promoter sequence can be operatively linked to a structural
CC	gene in the production of a recombinant vector. This vector may be
CC	used in the production of transformants. Recombinant genes bearing
CC	this promoters can be used to express many genes in cranial nerve
CC	cells. This is useful in the research of gene function and for
CC	research and therapy of brain disease.

```

XX      Sequence 1632 BP; 535 A; 343 C; 535 G; 219 T; 0 other;
SQ
Query Match          78.1%; Score 16.4; DB 14; Length 1632;
Best Local Similarity 94.4%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      4 cccgcgcccctcgtcgtc 21
         ||| ||||| |||||
DB      1530 cccgcgcccctcgtcgtc 1547

RESULT 11
AAQ39969
ID      AAQ39969 standard; DNA; 1632 BP.
XX
AC      AAQ39969;
XX
DT      28-JUL-1993 (first entry)
DE      PKC-gamma promoter region DNA.
XX
KW      Promoter; rat; protein kinase C; gamma; alpha; PKC; recombinant;
KM      vector; cranial; nerve cell; gene function; research; therapy;
KN      brain disease; upstream region; ss.
XX
OS      Rattus rattus.
XX
Key       Location/Qualifiers
FT       Promoter              1403..1632
         /*tag= a
         /note= "PKC-gamma promoter"
PM       JP05056781.A.
PD       09-MAR-1993.
PF       25-DEC-1990; 90JP-0405849.
PR       26-DEC-1989; 89JP-0334751.
PA       (TAKE ) TAKEDA CHEM IND LTD.
DR       WPI; 1993-120380/15.
XX
PT       New expression vector contg. protein kinase C promoter - useful
PT       for expressing genes in cranial nerve cell for treating related
PT       disease
XX
Discloure; Page 37-38; 39pp; Japanese.
CC       The sequences given in AAQ39968-69 represent the promoter and associated
CC       upstream regions isolated from the rat protein kinase C (PKC) gamma
CC       gene. The promoter sequence can be operatively linked to a structural
CC       gene in the production of a recombinant vector. This vector may be
CC       used in the production of transformants. Recombinant genes bearing
CC       this promoters can be used to express many genes in cranial nerve
CC       cells. This is useful in the research of gene function and for
CC       research and therapy of brain disease.
XX
SQ      Sequence 1632 BP; 535 A; 344 C; 534 G; 219 T; 0 other;
OY      Query Match          78.1%; Score 16.4; DB 14; Length 1632;
         Best Local Similarity 94.4%; Pred. No. 4.7e+02;
         Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
         4 cccgcgcccctcgtcgtc 21
         ||| ||||| |||||
DB      1530 cccgcgcccctcgtcgtc 1547

```

```
RESULT 12
AAK69214/c
ID AAK69214 standard; DNA; 115 BP.
XX
AC AAK69214;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24026.
XX
KM Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225417.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235370.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
```

PR	05-DEC-2000;	20000US-0251030.				
PR	05-DEC-2000;	20000US-0251988.				
PR	05-DEC-2000;	20000US-0256719.				
PR	06-DEC-2000;	20000US-0251479.				
PR	08-DEC-2000;	20000US-0251856.				
PR	08-DEC-2000;	20000US-0251868.				
PR	08-DEC-2000;	20000US-0251869.				
PR	08-DEC-2000;	20000US-0251989.				
PR	08-DEC-2000;	20000US-0251990.				
PR	11-DEC-2000;	20000US-0254097.				
PR	05-JAN-2001;	20010US-0259678.				
PA	(HUMA-) HUMAN GENOME SCI INC.					
PI	Rosen CA, Barash SC, Ruben SM;					
DR	WPI: 2001-483426/52.					
XX						
PT	Nucleic acids encoding human immune/haematopoietic antigen polypeptides,					
PR	useful for preventing, diagnosing and/or treating cancers and					
PT	metastasis -					
XX						
PS	Disclosure: SEQ ID NO 24026; 3071pp + Sequence Listing; English.					
XX						
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)					
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic					
CC	activity, and can be used in gene therapy and vaccine production. (I)					
CC	proteins and polynucleotides may be used in the prevention, diagnosis and					
CC	treatment of diseases associated with inappropriate (I) expression. For					
CC	example, they may be used to treat disorders associated with decreased					
CC	expression by rectifying mutations or deletions in a patient's genome					
CC	that affect the activity of (I) by expressing inactive proteins or to					
CC	supplement the patients own production of (I). Additionally, (I)					
CC	polynucleotides may be used to produce the secreted (I), by inserting					
CC	the nucleic acids into a host cell and culturing the cell to express the					
CC	protein. (I) proteins and polynucleotides may be used to prevent,					
CC	diagnose and treat immune/haematopoietic-related diseases, especially					
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703					
CC	to AAK67694 represent human immune/haematopoietic antigen genomic					
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169					
CC	represent sequences used in the exemplification of the present invention.					
XX						
SQ	Sequence 115 BP; 28 A; 21 C; 47 G; 19 T; 0 other;					
Query Match						
Best Local Similarity 77.1%; Score 16.2; DB 22; Length 115;						
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;						
QY	1	cctccgcgccttgcgtc 21				
DB	44	ccttccaccccttctcgtc 24				
RESULT 13						
ID	AAK69215/C					
XX	AAK69215 standard; DNA; 250 BP.					
XX	AAK69215;					
DT	06-NOV-2001 (first entry)					
XX						
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24027.					
XX						
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;					
XX	cytostatic; gene therapy; vaccine; metastasis; ds.					
OS	Homo sapiens.					
XX						
PN	WO200157182-A2.					
XX						
PD	09-AUG-2001.					
XX						

PR	17-JAN-2001	2001MO-US01354
PR	31-JAN-2000	2000US-0179065
PR	04-FEB-2000	2000US-0180628
PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0186350
PR	16-MAR-2000	2000US-0186874
PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0198129
PR	19-MAY-2000	2000US-0205512
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0214886
PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	11-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225457
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227189
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	14-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233299
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235636
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369

29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT

PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 24027; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (1)
CC amino acid sequences given in AAM82170 to AAM91921. (1) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (1)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (1) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (1) by expressing inactive proteins or to
CC supplement the patient's own production of (1). Additionally, (1)
CC polynucleotides may be used to produce the secreted (1), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 250 BP; 49 A; 59 C; 95 G; 46 T; 1 other;

Query Match 77.1%; Score 16.2; DB 22; Length 250;
Best Local Similarity 85.7%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cctcccgccctcgtcgtc 21
DB 179 CCTCCACCCTCTCTCTC 159
11111111111111111111

RESULT 14
AAF67072/c
ID AAF67072 standard; cDNA; 361 BP.
XX
XX AAF67072;
AC
XX
XX 09-APR-2001 (first entry)
DT
XX
XX Novel human polynucleotide, SEQ ID NO: 2828.
DE
XX
XX Human: cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200102568-A2.
PN
XX
XX 11-JAN-2001.
PD
XX
XX 30-JUN-2000; 2000WO-US18374.
PF
XX
XX 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEO INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klingner J, Kaasam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Cckenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
XX WPI; 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX

PS Claim 9; Page 965; 1046bp; English.
 XX
 CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.
 XX
 SQ Sequence 361 BP; 112 A; 68 C; 111 G; 70 T; 0 other;

Query Match 77.1%; Score 16.2; DB 22; Length 361;
 Best Local Similarity 85.7%; Pred. No. 6.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cctcccccgccttgctgc 21
 |||||
 DB 35 CCGCCCCGCCCTCTGCTGCTC 15

RESULT 15
 AAA31325/C
 ID AAA31325 standard; DNA; 377 BP.
 XX
 AC AAA31325;

DT 05-JUL-2000 (first entry)
 XX

DE Plant microsatellite marker #286.

XX Plant microsatellite sequence; core repeat sequence; detected; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;

KM variety identification; genetic variability evaluation; primer; ss.
 XX

OS Eucalyptus grandis.
 XX

XX WO967421-A1.
 PN

XX 29-DEC-1999.
 PD

XX 25-JUN-1999; 99MO-NZ00092.
 PF

XX 25-JUN-1998; 98US-0105307.
 PR

XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX

XX Havukkala JV, Bloksberg LN, Glenn M;
 PI

XX WPI; 2000-116958/10.
 DR

XX New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX

XX Claim 1; Page 161; 392pp; English.
 PS

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and

CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 XX
 SQ Sequence 377 BP; 72 A; 102 C; 130 G; 73 T; 0 other;

Query Match 77.1%; Score 16.2; DB 21; Length 377;
 Best Local Similarity 85.7%; Pred. No. 6.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cctcccccgccttgctgc 21
 |||||
 DB 354 CCGCCCCGCCCTCTGCTGCTC 334

Search completed: August 14, 2002, 22:06:37
 Job time: 11692 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:53:27 : Search time 203.42 Seconds
(without alignments)
25.358 Million cell updates/sec

Title: US-09-707-919-5

Perfect score: 21
Sequence: 1 cctccgcccttcgtcgtc 21

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 38533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/lna/5a COMB.seq:*
2: /cgn2_6/prodata/1/lna/5b COMB.seq:*
3: /cgn2_6/prodata/1/lna/5a COMB.seq:*
4: /cgn2_6/prodata/1/lna/5b COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCRD COMB.seq:*
6: /cgn2_6/prodata/1/lna/PCRD1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	355	4	US-09-043-303-1
2	21	100.0	623	4	US-09-043-303-5
3	21	100.0	4481	4	US-09-041-886-18
4	16.2	77.1	1743	3	US-09-032-365A-18
5	16.2	77.1	2316	6	5258283-6
6	16.2	77.1	9775	4	US-08-977-171-1
7	16.2	77.1	9934	4	US-08-977-171-2
8	15.2	72.4	673	6	5242798-4
9	15.2	72.4	1173	6	5242798-2
10	15.2	72.4	1838	5	PCT-US93-06251-85
11	15.2	72.4	2500	4	US-09-318-448-14
12	15.2	72.4	2542	1	US-08-120-960-1
13	15.2	72.4	4403765	4	US-09-103-840A-2
14	15	71.4	300	2	US-08-637-7598-27
15	15	71.4	300	4	US-08-871-355A-27
16	15	71.4	300	4	US-09-201-945-27
17	14.8	70.5	50	1	US-08-171-389-403
18	14.8	70.5	50	1	US-08-123-936-403
19	14.8	70.5	50	2	US-08-475-228A-403
20	14.8	70.5	50	2	US-08-482-080A-403
21	14.8	70.5	50	5	PCT-US93-12388-403
22	14.8	70.5	189	2	US-08-733-505A-51
23	14.8	70.5	189	2	US-08-733-505A-52
24	14.8	70.5	189	2	US-08-733-505A-53
25	14.8	70.5	189	2	US-08-733-505A-54
26	14.8	70.5	944	1	US-08-665-617-1
27	14.8	70.5	946	2	US-08-717-123-1

ALIGNMENTS

C 28	14.8	70.5	1105	3	US-08-985-335-2	Sequence 2, Appl
C 29	14.8	70.5	1105	4	US-09-410-372-2	Sequence 2, Appl
C 30	14.8	70.5	1256	4	US-09-318-448-42	Sequence 42, Appl
C 31	14.8	70.5	1273	4	US-09-318-448-45	Sequence 45, Appl
C 32	14.8	70.5	1275	4	US-09-318-448-41	Sequence 41, Appl
C 33	14.8	70.5	1462	4	US-09-434-288-4	Sequence 4, Appl
C 34	14.8	70.5	1908	4	US-09-318-448-36	Sequence 36, Appl
C 35	14.8	70.5	1945	1	US-08-724-194-1	Sequence 1, Appl
C 36	14.8	70.5	1958	4	US-09-401-476-1	Sequence 1, Appl
C 37	14.8	70.5	2095	4	US-09-401-476-3	Sequence 3, Appl
C 38	14.8	70.5	2678	1	US-08-724-194-2	Sequence 2, Appl
C 39	14.8	70.5	3283	4	US-09-061-709-8	Sequence 8, Appl
C 40	14.8	70.5	3412	4	US-09-061-709-6	Sequence 6, Appl
C 41	14.8	70.5	3791	3	US-09-377-310-1	Sequence 1, Appl
C 42	14.8	70.5	3963	1	US-08-464-961-1	Sequence 1, Appl
C 43	14.8	70.5	3963	3	US-08-907-800A-1	Sequence 1, Appl
C 44	14.8	70.5	3963	4	US-08-969-315-1	Sequence 1, Appl
C 45	14.8	70.5	3963	5	PCT-US96-08233-1	Sequence 1, Appl

RESULT 1
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043, 303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ. ID NOS: 17
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
; US-09-043-303-1

Query Match 100.0%; Score 21; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 0.69; 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0

Qy 1 cctccgcccttcgtcgtc 21
Db 158 cctccgcccttcgtcgtc 178

RESULT 2
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043, 303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ. ID NOS: 17

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 623
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (341)..(583)
FEATURE:
OTHER INFORMATION: Tsp-2
US-09-043-303-5

Query Match 100.0%; Score 21; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctcccgcccttgctgc 21
|||||
DB 158 cctcccgcccttgctgc 178

RESULT 3

US-09-041-886-18
Sequence 18, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Sharoz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

Query Match 100.0%; Score 21; DB 4; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cctcccgcccttgctgc 21

|||||
DB 460 CCTCCCGCCCTTGCTGTC 480

RESULT 4

US-09-032-365A-18/c
Sequence 18, Application US/09032365A
Patent No. 6114502
GENERAL INFORMATION:
APPLICANT: No. 6114502th, Michael
APPLICANT: Nishina, Patsy
APPLICANT: Naggart, Juergen
APPLICANT: No. 6114502en-1rauth, Konrad
TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
TITLE OF INVENTION: NEUROSENSORY DEFECTS
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,365A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-2C1P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650 327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1743 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-032-365A-18

Query Match 77.1%; Score 16.2; DB 3; Length 1743;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cctcccgcccttgctgc 21
|||||
DB 231 CCTCCCGCCCTTGCTGCC 211

RESULT 5

5258283-6
Patent No. 5258283
APPLICANT: FRAIER, MARVIN E.; MALLAVIA, LOUIS P.; SAMUEL,
JAMES E.; BACH, OSWALD G.
TITLE OF INVENTION: DETECTION AND DIFFERENTIATION OF COXIELLA
BURNETII IN BIOLOGICAL FLUIDS
NUMBER OF SEQUENCES: 17
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/425,856
FILING DATE: 23-OCT-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 927,779
FILING DATE: 05-NOV-1986
APPLICATION NUMBER: 795,207
FILING DATE: 05-NOV-1985
SEQ ID NO: 6
LENGTH: 2316
5258283-6

Query Match 77.1%; Score 16.2; DB 6; Length 2316;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cctcccgcccttcgtcgc 21
1 ||||| || |||||
DB 119 catcccgctcctcgtcgc 139

RESULT 6
US-08-977-171-1
Sequence 1, Application US/08977171
Patent No. 6232112
GENERAL INFORMATION:
APPLICANT: CATCHESIDE, DAVID E.
TITLE OF INVENTION: REAGENTS AND METHODS FOR DIVERSIFICATION
TITLE OF INVENTION: OF DNA
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 6232112west Center, 90 South 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,171
FILING DATE: 24-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SKOOG, MARK T
REGISTRATION NUMBER: 40,178
REFERENCE/DOCKET NUMBER: 10552.13US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9775 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-977-171-1

Query Match 77.1%; Score 16.2; DB 4; Length 9775;
Best Local Similarity 85.7%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cctcccgcccttcgtcgc 21
1 ||||| ||||| |||||
DB 699 cttccctcccttcgtcgc 719

RESULT 7
US-08-977-171-2
Sequence 2, Application US/08977171
Patent No. 6232112
GENERAL INFORMATION:
APPLICANT: CATCHESIDE, DAVID E.
TITLE OF INVENTION: REAGENTS AND METHODS FOR DIVERSIFICATION
TITLE OF INVENTION: OF DNA
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
STREET: 3100 No. 6232112west Center, 90 South 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,171
FILING DATE: 24-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SKOOG, MARK T
REGISTRATION NUMBER: 40,178
REFERENCE/DOCKET NUMBER: 10552.13US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 9934 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-977-171-2

Query Match 77.1%; Score 16.2; DB 4; Length 9934;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cctcccgcccttcgtcgc 21
1 ||||| ||||| |||||
DB 803 cttccctcccttcgtcgc 823

RESULT 8
5242798-4/C
APPLICANT: SUTCLIFFE, J. GERGOR
TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES CORRESPONDING
TO PORTIONS OF PROTEINIDS TRANSLATED FROM BRAIN-SPECIFIC MRNAS,
RECEPTORS, METHODS AND DIAGNOSTICS USING THE SAME
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/476,961
FILING DATE: 07-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 58,620
FILING DATE: 03-JUN-1987
APPLICATION NUMBER: 516,136
FILING DATE: 21-JUL-1983
SEQ ID NO: 4:

LENGTH: 673
5242798-4

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 6; Length 673;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cctcccgcccttcgtcgt 20
Db 327 CCTCCCGCCCGCCTCCTCCT 308

RESULT 9
5242798-2/c
PATENT NO. 5242798
APPLICANT: SUTCLIFFE, J. GERGOR
TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES CORRESPONDING
TO PORTIONS OF PROTEINOIDS TRANSLATED FROM BRAIN-SPECIFIC MRNAS,
RECEPTORS, METHODS AND DIAGNOSTICS USING THE SAME
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/476,961
FILING DATE: 07-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 58,620
FILING DATE: 03-JUN-1987
APPLICATION NUMBER: 516,136
FILING DATE: 21-JUL-1983
SEQ ID NO: 2;
LENGTH: 1173
5242798-2

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 6; Length 1173;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cctcccgcccttcgtcgt 20
Db 814 CCTCCCGCCCGCCTCCTCCT 795

RESULT 10
PCT-US93-06251-85
Sequence 85, Application PC/TUS9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343

TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 1838 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-85

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 5; Length 1838;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 cctcccgcccttcgtcgt 20
Db 323 CCTCCCGCCCGCCTCCTCCT 342

RESULT 11
US-09-318-448-14/c
Sequence 14, Application US/09318448
PATENT NO. 6210950
GENERAL INFORMATION:
APPLICANT: Johnson, William G.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 2500
TYPE: DNA
ORGANISM: Homo sapiens
US-09-318-448-14

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 4; Length 2500;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cctcccgcccttcgtcgt 21
Db 2163 CTCCCGCCCGCCTTCTGAC 2144

RESULT 12
US-08-120-960-1/c
Sequence 1, Application US/08120960
PATENT NO. 553225
GENERAL INFORMATION:
APPLICANT: KRAUS, JAN P
TITLE OF INVENTION: DNA SEQUENCE ENCODING HUMAN
TITLE OF INVENTION: CYSTATINONINE B-SYNTHASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 4350 LA JOLLA VILLAGE DRIVE, SUITE 300
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/120,960
; FILING DATE: 12-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PEPPER PH.D., FREDERICK W.
; REGISTRATION NUMBER: 31,286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4410
; TELEFAX: 619-453-2839
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: mat.peptide
; LOCATION: 181..1834
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 181..1834
; US-08-120-960-1
```

```

Query Match          72.4%; Score 15.2; DB 1; Length 2542;
Best Local Similarity 85.0%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      2 ctccgcgcgccttcgtc 21
        |||||
Db      2185 CTCGCCGCCCTCTCTGAC 2166
```

```

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```

Query Match          72.4%; Score 15.2; DB 4; Length 4403765;
Best Local Similarity 85.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      2 ctccgcgcgccttcgtc 21
        |||||
Db      2249890 CTCGACGCCCTCTGCGTC 2249871
```

```

RESULT 14
US-08-637-759B-27/c
; Sequence 27, Application US/08637759B
```

```

; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; US-08-637-759B-27
```

```

Query Match          71.4%; Score 15; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 ctccgcgcgccttcg 16
        |||||
Db      194 CTCGCCGCCCTCTG 180
```

```

RESULT 15
US-08-871-355A-27/c
; Sequence 27, Application US/08871355A
; Patent No. 601569
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPLS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Partial sequence of *Salmonella typhimurium*
ORGANISM: virulence gene
US-08-871-355A-27

Query Match 71.4%; Score 15; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ctccccgccctcg 16
|||||
Db 194 CTCGCCGCCCTTCG 180

Search completed: August 14, 2002, 21:55:21
Job time: 13754 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:17 ; Search time 7749.14 Seconds

(Without alignments)
36.576 Million cell updates/sec

Title: US-09-707-919-5

Sequence: 1 cctcccgcctcctcgcgc 21

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estha:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	482	9	AL039573 DKFZP434D
2	21	100.0	500	10	B1547486
3	21	100.0	1100	10	BM455214 AGENCOURT
4	20	95.2	364	10	BE457923
5	18	85.7	458	9	AW785621
6	17.8	84.8	418	9	AU058001
7	17.8	84.8	563	10	B192379
8	17.8	84.8	637	10	BG540583
9	17.8	84.8	722	12	AG064522
10	17.8	84.8	782	10	BE972945
11	17.8	84.8	808	3	B1645027
12	17.8	84.8	846	12	CNS03W7C
13	17.8	84.8	853	10	B1958145
14	17.8	84.8	871	10	BM415654
15	17.8	84.8	884	9	AL535465
16	17.8	84.8	1036	12	CNS036KK
17	17.8	84.8	1078	10	BM453087

18	17.8	84.8	1167	10	BM455134	BM455134 AGENCOURT
19	17.4	82.9	288	10	B1802666	B1802666 H086B11.E
20	17.4	82.9	425	10	BM489609	BM489609 pgm2n.pk0
21	17.4	82.9	440	10	BF073435	BF073435 224498.MA
22	17.4	82.9	442	12	AQ321683	AQ321683 RPT111-96
23	17.4	82.9	633	12	AG185635	AG185635 Pan trol1
24	17.4	82.9	921	12	CNS005K2	AL059599 Drosophil
25	17.4	82.9	1348	10	BE727818	BE727818 601564188
26	17.2	81.9	579	12	CNS02KMG	AL202057 Tetradon
27	17	81.0	990	10	BF245188	BF245188 601863841
28	16.8	80.0	276	9	AW762952	AW762952 urf8b08.Y
29	16.8	80.0	298	9	BG238247	BG238247 sab59b09.Y
30	16.8	80.0	302	10	BG788854	AG074168 RPT-23-4
31	16.8	80.0	313	12	AZ074168	AG074168 RPT-23-4
32	16.8	80.0	415	12	BH616904	BH616904 SALK_0356
33	16.8	80.0	443	9	A1443769	A1443769 sa29g04.Y
34	16.8	80.0	444	12	BH018157	BH018157 L130K.d.H
35	16.8	80.0	456	9	AW211389	AW211389 u080G08.Y
36	16.8	80.0	459	9	A1966080	A1966080 sc27a08.Y
37	16.8	80.0	463	12	BH018155	BH018155 L130b.d.H
38	16.8	80.0	471	10	B1943629	B1943629 sr24c09.Y
39	16.8	80.0	473	9	A1780467	A1780467 EST261346
40	16.8	80.0	476	12	BH172109	BH172109 SALK_0052
41	16.8	80.0	482	10	B1396634	B1396634 ro59f07.Y
42	16.8	80.0	482	10	B1944789	B1944789 sa66c11.Y
43	16.8	80.0	506	12	BH169336	BH169336 SALK_0010
44	16.8	80.0	525	10	B1802432	B1802432 H079G10.E
45	16.8	80.0	535	9	AJ241138	AJ241138 AJ241138

ALIGNMENTS

RESULT 1	AL039573	482 bp	mRNA	linear	EST 29-FEB-2000
LOCUS	DKFZP434D1311	r1 434	(synonym: htes3)	Homo sapiens	CDNA clone
DEFINITION	DKFZP434D1311 5', mRNA sequence.				
ACCESSION	AL039573				
VERSION	AL039573.1	GI:5408612			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann				
TITLE	EST (Duesterhoeft, et al.)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Duesterhoeft A				

Am Kiofepseplitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.

No si sequence available.
This clone (DKFZP434D1311) is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@r2pd.de.

FEATURES

source
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP434D1311"
/clone.lib="434 (synonym: htes3)"
/tissue.type="testis"
/dev.stage="adult"
/lab.note="DH10B"
/note="Vector: pSport1; site_1: NotI; site_2: SalI"
BASE COUNT 49 a 218 c 145 g 70 t

ORIGIN

Query Match 100.0%; Score 21; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctccgcccttcgtcgc 21
|||||

Db 107 CCTCCCCGCCCTTCGTCGTC 127

RESULT 2
B1547486 500 bp mRNA linear EST 05-SEP-2001
LOCUS B1547486
DEFINITION 603191091F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:5262335 5',
mRNA sequence.
ACCESSION B1547486
VERSION B1547486.1 GI:15434798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 500)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11661 row: e column: 24
High quality sequence stop: 485.
Location/Qualifiers
1..500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5262335"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: p Bluescript (modified
p Bluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gtcgcg
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 57 a 222 c 150 g 71 t

ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctccgcccttcgtcgc 21
|||||

Db 110 CCTCCCCGCCCTTCGTCGTC 130

RESULT 3
BMA55214 1100 bp mRNA linear EST 05-FEB-2002
LOCUS BMA55214

DEFINITION AGENCOURT 6405612 NIH_MGC_85 Homo sapiens CDNA clone IMAGE:5500163
5', mRNA sequence.
ACCESSION BMA55214
VERSION BMA55214.1 GI:18504254
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1100)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lou Straud
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12134 row: k column: 12
High quality sequence stop: 623.
Location/Qualifiers
1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 240 a 329 c 306 g 219 t 6 others

ORIGIN

Query Match 100.0%; Score 21; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctccgcccttcgtcgc 21
|||||

Db 81 CCTCCCCGCCCTTCGTCGTC 101

RESULT 4
BE457923 364 bp mRNA linear EST 26-JUL-2000
LOCUS BE457923
DEFINITION us99c12.x1 Soares-lymus_2NDMT Mus musculus CDNA clone
IMAGE:3326518 3' similar to TR:070305 070305 SPINOCEREBELLAR ATAXIA
2 HOMOLOG ;, mRNA sequence.
ACCESSION BE457923
VERSION BE457923.1 GI:9480561
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 364)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1070682

FEATURES
source Possible reversed clone: polyT not found.
Location/Qualifiers
1. 364
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3326518"
/clone_lib="Soares_thymus_2nbmr"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
3'; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 51 a 126 c 173 g 14 t

ORIGIN

Query Match 95.2%; Score 20; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgccctcgtcgt 20
|||||

Db 338 CCTCCCCGCCCTTCGTCT 319

RESULT 5
AW785621 458 bp mRNA linear EST 09-JUL-2000
DEFINITION AW785621 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW785621
VERSION AW785621.1 GI:7842334
KEYWORDS EST.
SOURCE
ORGANISM Sus scrofa
Pig.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 458)
Fahrenkrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and a/c-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 51 row: P column: 5
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 458
/organism="Sus scrofa"
/db_xref="taxon:9623"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"

FEATURES
source

/lab_host="DH10B"
/note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 85 a 128 c 114 g 131 t

ORIGIN

Query Match 85.7%; Score 18; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 6.6e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cctcccgccctcgtcgt 18
|||||

Db 354 CCTCCCCGCCCTTCGTC 371

RESULT 6
AU058001 418 bp mRNA linear EST 29-APR-1999
DEFINITION AU058001 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
clone S21981_1A, mRNA sequence.
ACCESSION AU058001
VERSION AU058001.1 GI:4716885
KEYWORDS EST.
SOURCE
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 418)
Yamamoto,K. and Sasaki,T.
Rice cDNA from mature leaf
Unpublished (1999)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@agr.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/
PROJECT "RGP".

FEATURES
source Location/Qualifiers
1. 418
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S21981_1A"
/clone_lib="Oryza sativa mature leaf Nipponbare"
/tissue_type="mature leaf"

BASE COUNT 167 a 83 c 79 g 88 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 418;
Best Local Similarity 90.5%; Pred. No. 7.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 cctcccgccctcgtcgt 21
|||||

Db 50 CCTTCCCTCCCTTCGTCGTC 70

RESULT 7
BJ192379 563 bp mRNA linear EST 24-JAN-2002
DEFINITION BJ192379 normalized full length cDNA library, chloronemata,
caulonemata and rhizoid-like protonemata Physcomitrella patens
subsp. Patens cDNA clone pphnism20 5', mRNA sequence.
ACCESSION BJ192379
VERSION BJ192379.1 GI:18360317
KEYWORDS EST.
SOURCE Physcomitrella patens subsp. patens.

ORGANISM Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 563)
AUTHORS Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,
Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe
,M.
TITLE Comparison of the moss Physcomitrella patens genome with flowering
plants genome
JOURNAL unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yatai Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
A backbone of the vector is pBluescript II, that was in vivo
excised from a modified lps phage vector (Mo bi Rec, Germany). XhoI
digested-5' end of cDNA is ligated to SalI site of the vector, and
the BamHI digested-3' end including poly-A tail is ligated to BamHI
site of the vector. cDNA insert could be amplified with
conventional T7 and T3 primers. This normalized full-length cDNA
library was generated basically according to the method described
in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
Protonemata were blended by the POLYTRON, and then cultivated on
the BCD medium containing 1um NAA (naphthalene acetic acid) for 8
to 11 days under the continuous light.
Location/Qualifiers
source 1..563
/organism="Physcomitrella patens subsp. patens"
/db_xref="taxon:145481"
/clone="pphnl5m20"
/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and rhizoid-like protonemata"
/tissue_type="mixture of chloronemata, caulonemata and
rhizoid-like protonemata"
BASE COUNT 75 a 167 c 137 g 184 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 563;
Best Local Similarity 90.5%; Pred. No. 7.8e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cctcccgcccttcgtc 21
||||||| ||||| |||||
Db 136 CCTCCACCCCTCGTCGTC 156

RESULT 8
BG540583 637 bp mRNA linear EST 03-APR-2001
LOCUS 602370437F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4 94889 5',
DEFINITION mRNA sequence.
ACCESSION BG540583
VERSION BG540583.1 GI:13532816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 637)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: CLOUTech Laboratories, Inc.
cDNA Library Preparation: CLOUTech Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L16M1519 row: j column: 10
High quality sequence stop: 218.
Location/Qualifiers
source 1..637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4694889"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1;
SfiI (ggccgctcgcc); Site 2: SfiI (ggccatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTAATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCGCCGATG-dT(30)NN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT 194 a 110 c 278 g 55 t
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 637;
Best Local Similarity 90.5%; Pred. No. 7.8e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cctcccgcccttcgtc 21
||||||| ||||| |||||
Db 617 CCTCCCGCCCGCTCGTCGTC 597

RESULT 9
AG064522 727 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-053K01.R, genomic survey sequence.
DEFINITION AG064522
ACCESSION AG064522.1 GI:16616324
VERSION GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-053K01.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 727)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submissiion
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC)
1-7-22 Suehiro-chou,Tsukumi-ku, Tokohama, Kanagawa 230-0045, Japan
(E-mail:chimpanzee@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
source 1..727
/organism="Pan troglodytes"
/db_xref="taxon:9598"

/clone="PTB-053K01.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 BASE COUNT 120 a 334 c 91 g 181 t 1 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 12; Length 727;
 Best Local Similarity 90.5%; Pred. No. 7.9e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 cctcccgcccttcgtcgc 21
 |||||||
 Db 387 CCTCCCGCCCTTCTCTC 407

RESULT 10
 BE972945/c 782 bp mRNA linear EST 04-OCT-2000
 LOCUS 601652591R2 NIH_MGC_82 Homo sapiens CDNA clone IMAGE:3935638 3',
 DEFINITION mRNA sequence.

ACCESSION BE972945
 KEYWORDS BE972945.1 GI:10586281
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 782)
 NIH-MGC http://mhc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 plate: LNCM778 row: f column: 23
 High quality sequence stop: 3.
 Location/Qualifiers

FEATURES
 Source 1..782
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_82"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: testis; Vector: pBNR-LIB (Clontech); Site: 1:
 SfilI (ggcgccctgcgc); Site 2: SfilI (ggcattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTAATGGC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGAGCGCGCGACAGT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

BASE COUNT 195 a 173 c 274 g 138 t 2 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 782;
 Best Local Similarity 90.5%; Pred. No. 7.9e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 cctcccgcccttcgtcgc 21
 |||||||
 Db 200 CCTCCCGCCCTTCTCTC 180

RESULT 11
 B1645027 standard; RNA; EST; 808 BP.
 ID B1645027
 AC B1645027;
 XX
 SV B1645027.1
 XX
 XX
 DT 13-SEP-2001 (Rel. 69, Created)
 DT 13-SEP-2001 (Rel. 69, Last updated, Version 1)
 XX
 XX

DE OP2828 Mixed Stage EST's from Globodera pallida, the potato cyst nematode
 DE Globodera pallida cDNA, mRNA sequence.
 DE
 DE EST.
 XX
 XX
 OS Globodera pallida
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.

XX [1]
 RP 1-808
 RA Heer J., Sosinski B., Pokrzywa R.M., Wary A., Opperman C.;
 RT "Mixed stage EST's from Globodera pallida, the potato cyst nematode";
 RL Unpublished.
 XX

CC Contact: Opperman, C
 CC Center for the Biology of Nematode Parasitism
 CC NC State University; IACR-Rothamsted
 CC Campus Box 7616; Raleigh, NC 27695, USA
 CC Tel: 919.515.6699
 CC Fax: 919.515.9500
 CC Email: warchogunity.ncsu.edu
 CC No Homology found. ; GT11-6PCN_F_F01_PCN_6_F_011.ab1.seq.screen.
 CC

XX Key Location/Qualifiers

XX FH 1..808
 XX FT /db_xref="taxon:36090"
 FT /note="Vector: lambda GT11; This is a collaborative effort
 FT between IACR-Rothamsted and North Carolina State
 FT University. The library was constructed from mixed stage G.
 FT pallida in lambda GT11 by Paul Burroughs,
 FT IACR-Rothamsted."
 FT /organism="Globodera pallida"
 FT /clone_lib="Mixed Stage EST's from Globodera pallida, the
 FT potato cyst nematode"
 FT

XX SQ Sequence 808 BP; 106 A; 240 C; 163 G; 297 T; 2 other;

Query Match 84.8%; Score 17.8; DB 3; Length 808;
 Best Local Similarity 90.5%; Pred. No. 7.9e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 cctcccgcccttcgtcgc 21
 |||||||
 Db 769 CCTCCCGCCCTTCTCTC 789

RESULT 12
 CNS03W7G 846 bp DNA linear GSS 18-MAY-2000
 LOCUS CNS03W7G/c
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 063N01 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL263365.1 GI:7985024
 VERSION AL263365
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS
1 (bases 1 to 846)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 846)
Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL
REFERENCE
AUTHORS
3 (bases 1 to 846)
Genoscope.
TITLE
Direct Submission
JOURNAL
COMMENT
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1..846
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="063N01"
/clone_1lb="G"
/note="Genoscope sequence ID : C08G063G01LP1-end : T7"
BASE COUNT
240 a 197 c 199 g 196 t 14 others
ORIGIN

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 12; Length 846;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cctcccgcccttcgtcgc 21
|||||
Db 396 CCTCCCGCCCTCCTCGCGTC 376

RESULT 13
BI958145/C 853 bp mRNA linear EST 22-OCT-2001
LOCUS
DEFINITION
HVSME0013J18f Hordeum vulgare rachis EST library HVCNDA0015 (normal) Hordeum vulgare cDNA clone HVSME0013J18f, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
barley.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.
1 (bases 1 to 853)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons J., Oates,R. and Main,D.
TITLE
Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL
COMMENT
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 459

Seq primer: ATTATACCTCACTAAGG
High quality sequence stop: 472.
Location/Qualifiers
1..853
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSME0013J18f"
/clone_1lb="Hordeum vulgare rachis EST library HVCNDA0015 (normal)"
/tissue_type="Rachis"
/lab_host="TJC121"
/note="Vector: pBluescript SK(-). Site 1: EcoRI; Site 2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinohfs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinohfs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9ppages/bgn/31/cover.html>)"

BASE COUNT
138 a 266 c 331 g 117 t 1 others
ORIGIN

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 10; Length 853;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cctcccgcccttcgtcgc 21
|||||
Db 692 CCTCCCGCCCTCCTCCTC 672

RESULT 14
BM415654 871 bp mRNA linear EST 28-JAN-2002
LOCUS
DEFINITION
OP20732 Mixed Stage EST's from Globodera pallida, the potato cyst nematode Globodera pallida cDNA, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Globodera pallida.
Globodera pallida
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
1 (bases 1 to 871)
Heer,J., Sosinski,B., Pokrzywa,R.M., Warry,A. and Opperman,C.
TITLE
Mixed Stage EST's from Globodera pallida, the potato cyst nematode Unpublished (2001)
JOURNAL
COMMENT
Contact: Opperman, C
Center for the Biology of Nematode Parasitism
NC State University; IACR-Rothamsted
Campus Box 7616; Raleigh, NC 27695, USA
Tel: 919.515.6699
Fax: 919.515.9500
Email: warthog@unity.ncsu.edu
GT11-6PCN_F_F01_PCN_6_F_011.abl.
Location/Qualifiers

FEATURES

source

1. 871
 /organism="Globodera pallida"
 /db_xref="taxon:36090"
 /clone_lib="Mixed Stage EST's from Globodera pallida, the potato cyst nematode"
 /note="Vector: lambda GT11; This is a collaborative effort between IACR-Rothamsted and North Carolina State University. The library was constructed from mixed stage G. pallida in lambda GT11 by Paul Burroughs, IACR-Rothamsted."

BASE COUNT 143 a 248 c 172 g 304 t 4 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 871;
 Best Local Similarity 90.5%; Pred. No. 7.9e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cctcccgcgccttcgtcgc 21
 ||||| ||||| ||||| |||||
 Db 832 CCMCCCGCTCCCTCGCCGTC 852

RESULT 15
 AL535465 884 bp mRNA linear EST 13-FEB-2001
 LOCUS AL535465 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF009YM04 5
 DEFINITION prime, mRNA sequence.

ACCESSION AL535465
 VERSION AL535465.1 GI:12798958
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 884)
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1. 884
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF009YM04"
 /clone_lib="LTI_FL013_FBrn1"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: PCWVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifeleth.com URL : http://fulllength.invitrogen.com"

BASE COUNT 195 a 176 c 320 g 189 t 4 others
 ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 884;
 Best Local Similarity 90.5%; Pred. No. 7.9e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cctcccgcgccttcgtcgc 21
 ||||| ||||| ||||| |||||
 Db 90 CCFACCGCCCTTCGCGCTC 70

Biochemical Technology, Delhi university Campus, Mall Road, Delhi									
110 007, India									
location/Qualifiers									
1..264									
/organism="Papio hamadryas"									
/db_xref="taxon:9557"									
<1..>264									
/gene="SCA2"									
/note="splnocerebellar ataxia 2"									
BASE COUNT									
25 a 130 c 78 g 31 t									
ORIGIN									
Query Match									
Best Local Similarity 100.0%; Score 32; DB 9; Length 264;									
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
OY 1 cgcgaaccgcgcctcccgctcgagccgc 32									
Db 73 CGCCAAACCGCGCTCCCGCTCGGCGCCG 104									
RESULT 2									
AF330028 390 bp DNA linear PRI 08-NOV-2001									
LOCUS									
DEFINITION Pan troglodytes SCA2 gene, partial sequence.									
ACCESSION AF330028									
VERSION AF330028.1 GI:12382830									
KEYWORDS									
SOURCE									
ORGANISM									
Pan troglodytes									
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.									
1 (bases 1 to 390)									
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and									
Brahmachari,S.K.									
CAG repeat instability at SCA2 locus: anchoring CAA interruptions									
and linked single nucleotide polymorphisms									
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)									
11689490									
2 (bases 1 to 390)									
Choudhry,S. and Brahmachari,S.K.									
Direct Submision									
Submitted (21-DEC-2000) Functional Genomics Unit, Center for									
Biochemical Technology, Delhi University Campus, Mall Road, Delhi									
110 007, India									
Location/Qualifiers									
1..390									
/organism="Pan troglodytes"									
/db_xref="taxon:9598"									
1..390									
/note="microsatellite"									
/rpt_type=tandem									
/rpt_unit=cag									
<1..>390									
/gene="SCA2"									
/note="splnocerebellar ataxia 2"									
BASE COUNT									
48 a 183 c 110 g 49 t									
ORIGIN									
Query Match									
Best Local Similarity 100.0%; Score 32; DB 9; Length 390;									
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
OY 1 cgcgaaccgcgcctcccgctcgagccgc 32									
Db 67 CGCCAAACCGCGCTCCCGCTCGGCGCCG 98									
RESULT 3									
AC004085/c 231758 bp DNA linear HTG 06-NOV-2000									

DEFINITION	ACCSSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered pieces	AC004085	AC004085.6	GI:11079383	HTGS, PHASE1: HTGS, DRAFT, human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
						1 (bases 1 to 231758)	Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Blinagle,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christophoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Den,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M., Falls,T., Ferragudo,D., Flagg,N., Ford,J., Foster,P., Frotz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunnathne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Huiyk,S., Hune,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyvt,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louisedge,M., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawliny,E., Mcleod,M.P., Meador,M., Mel,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S., Ogun,M., Okunolu,G., Orlunge,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Plokers,R., Prlmus,E., Pu,L.L., Quilis,M., Ren,Y., Rives,M., Rojas,A., Rojodokan,I., Rolfe,M., Ruiz,S., Seaver,J., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabori,P., Tameis,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.
						Unpublished	Direct Submission
						2 (bases 1 to 231758)	Worley,K.C.
						Direct Submission	Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
						On Nov 3, 2000 this sequence version replaced	gi:9966929.
						Genome Center	
						Center: Baylor College of Medicine	
						Center code: BCM	
						Web site: http://www.hgsc.bcm.tmc.edu/	
						Contact: hgsc-help@bcm.tmc.edu	
						Project Information	
						Center project name: UG	
						Center clone name: RP11-42B1	
						Summary Statistics	
						Assembly program: Phrap; version 0.990329	
						Consensus quality: 224768 bases at least Q40	
						Consensus quality: 229074 bases at least Q30	
						Consensus quality: 230948 bases at least Q20	
						Estimated insert size: 227237, sum-of-contigs estimation	
						Estimated insert size: 317511, agatose-tp estimation	
						Quality coverage: 6.3x in Q20 bases; agatose-tp estimation	

Quality coverage: 8.8x in Q20 bases: sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 33241: contig of 33241 bp in length
33242 33341: gap of unknown length
33342 56391: contig of 23050 bp in length
56392 56491: gap of unknown length
56492 81323: contig of 24833 bp in length
81324 81423: gap of unknown length
81424 102538: contig of 21115 bp in length
102539 119710: gap of unknown length
119711 119810: contig of 17072 bp in length
119811 136913: gap of unknown length
136914 137013: gap of unknown length
137014 153285: contig of 16272 bp in length
153286 153385: gap of unknown length
153386 167987: contig of 14602 bp in length
167988 168087: gap of unknown length
168088 178731: contig of 10644 bp in length
178732 178831: gap of unknown length
178832 186641: contig of 7810 bp in length
186642 186741: gap of unknown length
186742 193215: contig of 6474 bp in length
193216 193315: gap of unknown length
193316 201310: contig of 7995 bp in length
201311 201410: gap of unknown length
201411 208647: contig of 7237 bp in length
208648 208747: gap of unknown length
208748 213802: contig of 5055 bp in length
213803 213902: gap of unknown length
213903 218049: contig of 4147 bp in length
218050 218149: gap of unknown length
218150 223316: contig of 5167 bp in length
223317 223416: gap of unknown length
223417 227389: contig of 3973 bp in length
227390 227489: gap of unknown length
227490 229032: contig of 1543 bp in length
229033 229132: gap of unknown length
229133 230651: contig of 1519 bp in length
230652 230751: gap of unknown length
230752 231758: contig of 1007 bp in length.

FEATURES
Source
1. .231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

Query Match 100.0%; Score 32; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 89265 CGCACAACCGCGCTCCCGCTCGGCGCCG 89234

RESULT 4 ARI59544 355 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59544
DEFINITION Sequence 1 from patent US 6251589.

ACCESSION ARI59544
VERSION ARI59544.1 GI:16222225

KEYWORDS
SOURCE
ORGANISM
Unknown.

REFERENCE
1 (bases 1 to 355)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
Patent: US 6251589-A 1 26-JUN-2001;

JOURNAL
Location/Qualifiers
1. .355

BASE COUNT 20 a 176 c 102 g 55 t 2 others
ORIGIN

Query Match 98.8%; Score 31.6; DB 6; Length 355;
Best Local Similarity 96.9%; Pred. No. 5.9;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 219 CGCACAACCGCGCTCCCGCTCGGCGCCG 250

RESULT 5 ARI59558 572 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59558
DEFINITION Sequence 18 from patent US 6251589.

ACCESSION ARI59558
VERSION ARI59558.1 GI:16222251
KEYWORDS
SOURCE
ORGANISM
Unknown.

REFERENCE
1 (bases 1 to 572)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
Patent: US 6251589-A 18 26-JUN-2001;

JOURNAL
Location/Qualifiers
1. .572

BASE COUNT 34 a 277 c 174 g 85 t 2 others
ORIGIN

Query Match 98.8%; Score 31.6; DB 6; Length 572;
Best Local Similarity 96.9%; Pred. No. 5.2;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 219 CGCACAACCGCGCTCCCGCTCGGCGCCG 250

RESULT 6 ARI59546 623 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59546
DEFINITION Sequence 5 from patent US 6251589.

ACCESSION ARI59546
VERSION ARI59546.1 GI:16222229
KEYWORDS
SOURCE
ORGANISM
Unknown.

REFERENCE
1 (bases 1 to 623)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
Patent: US 6251589-A 5 26-JUN-2001;

JOURNAL
Location/Qualifiers
1. .623

REFERENCE 2 (bases 1 to 322)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
source 1..322
/organism="Macaca radiata"
/db_xref="taxon:9548"
gene <1..>322
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 32 a 155 c 95 g 40 t
ORIGIN

Query Match 95.0%; Score 30.4; DB 9; Length 322;
Best Local Similarity 96.9%; Pred. No. 13;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cgcgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 96 CGCAACCGCGCTCCGCTCGCGCCGC 127

RESULT 13
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS Presbytlis entellus SCA2 gene, partial sequence.
DEFINITION AF330030
ACCESSION AF330030.1 GI:12382832
VERSION
KEYWORDS
SOURCE Human langur.
ORGANISM Presbytlis entellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Colobinae; Presbytis.
1 (bases 1 to 384)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
REFERENCE 2 (bases 1 to 384)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
source 1..384
/organism="Presbytis entellus"
/db_xref="taxon:9574"
gene <1..>384
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 46 a 178 c 109 g 51 t
ORIGIN

Query Match 95.0%; Score 30.4; DB 9; Length 384;
Best Local Similarity 96.9%; Pred. No. 13;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cgcgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 73 CGCAACCGCGCTCCGCTCGCGCCGC 104

RESULT 14
AF330029

LOCUS AF330029 409 bp DNA linear PRI 08-NOV-2001
DEFINITION Gorilla gorilla SCA2 gene, partial sequence.
ACCESSION AF330029
VERSION AF330029.1 GI:12382831
KEYWORDS
SOURCE
ORGANISM gorilla.
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Gorilla.
1 (bases 1 to 409)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
REFERENCE 2 (bases 1 to 409)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
source 1..409
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
gene <1..>409
/gene="SCA2"
/note="spinocerebellar ataxia 2"
BASE COUNT 35 a 196 c 120 g 58 t
ORIGIN

Query Match 91.9%; Score 29.4; DB 9; Length 409;
Best Local Similarity 96.8%; Pred. No. 25;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gccgaaccgcgcctcccgctgcgcgcgcgc 32
|||||
Db 102 GCCAACC CGCGCTCCGCTCGCGCCGC 132

RESULT 15
AF041472 4225 bp mRNA linear ROD 28-NOV-2001
LOCUS Mus musculus ataxin-2 (SCA2) mRNA, complete cds.
DEFINITION AF041472
ACCESSION AF041472.1 GI:3005019
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4225)
Nechiporuk,T.T., Huynh,D.P., Figueroa,K., Sabha,S., Nechiporuk,A.V.
and Pulst,S.M.
The mouse SCA2 gene: cDNA sequence, alternative splicing and
protein expression
Hum. Mol. Genet. 7 (8), 1301-1309 (1998)
PUBMED 9668173
REFERENCE 2 (bases 1 to 4225)
AUTHORS Nechiporuk,T.T., Figueroa,K., Sabha,S., Nechiporuk,A.V. and
Pulst,S.M.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Medicine/Neurology, Cedars-Sinai Medical
Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
FEATURES
source 1..4225
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="12"
/map="12q23.1"

gene
CDS

```
1..4225
/gene="SCA2"
27..3884
/gene="SCA2"
/codon_start=1
/product="ataxin-2"
/db_xref="GI:3005020"
/translation="MRSSTAAVQRPACGDPPEPRRPAQMARRSLPRTANRGGRGAVA
VPSAGPPRGPGAPPRGPPSCASDCFCNGHGASRPCSRRLGVCGRPPEVVL
ALAPATPARACPGVRASPPRGVSSSARPAAGCPACEPVYGLTMSLPOPOP
APATGRKPGGGLSSPGAAPASAAVTSASVYPAPAPVSSAAAGGRRGGRNS
SKGLPQPTISFDGITANVAMHILTSVGSKECYOKNGGITEGVFKTSPKCDLYD
AAHEKSTESSGPKREELIESVLFKCSDFVVQFKDTSIARDAFTDSALSAKVG
EHKEKDLPEWDAGELTASELELENDVSNMGDPNDMFYNEENYCVSTYDSSLST
VPLERNSEEFLEKREARANOLAEIIESSAOYKARVALENDDRSE. EKYTAVORNCSDR
EGGPMTRDNKYIPQORNRREVLNMGSGROSSPRMGQPGGSMPLRAAHTHRPSPRPSRPS
GSDQRYVNGVPMPPSPCSPSSSRPPSRVOSGPNLPPRAATHRPSPRPSRPS
HPSAHGSPAPVSTMTRKMSSEGRPMKSPRAORHPRNHRVSAAGCSMSGLFEVSHNP
SEAAAPPVARTSPAGTWSVSGVFRLSPTKTHRRSPKQSSIGNSPGPVLASPOAG
IIPAEAVSMPPVPAASPTPASPASNRLTPSIEAKDSRIODORONSPAGSKENVKASET
SPFSKADNKGMSPVYSEHRKOIDILKKFNDFRLQPSSTSEMDQLLSKNREGESR
DLIKDTEASAKDSFIDSSSSSNCSTSGSSTNSPSISPSMLSNAEHRGPEVTSQGV
OTSSPACKEKNDREKKDTEOVREKSTLNPNAKEPNRPSFOPKPTPTSPRPOAO
PSPSMVGHOOAPAPVYTOPYCFAPNMMPVYVSPGVQPLPIPMTPMVPNOAKTYRAGK
VPMPOORODHOSHTMHPASAGPPIVATPPAYSTQYVAISPOQFPNOPLVQHYPH
YOSQHPHYSPVIOGNARMAPPAAHQPLVSSSAQFGAHEOTHMTACPKLPTMKE
TSPSEFYAISTGSLAOQVAHPNALHPHTPHPOPSATPTGQOOSOHGSHPAVSPVQH
HOHOAAOLHLASPOQOSAIVHAGLAPTPPSMTPASNTQPOSSFPAAQOTVFTIHP
HYOPAYTTPPHMAHYPOAHVOSGMVPSHPTAHAPMMLMTTOPPGKALALQSLQPI
VSTTAHPPTMHPSPVOAHHQOOL"
```

BASE COUNT 1007 a 1324 c 1042 g 851 t 1 others
ORIGIN

Query Match 75.0%; Score 24; DB 10; Length 4225;
Best Local Similarity 84.4%; Pred. No. 5.2e+02;
Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 cgcacaaccgcgcctcccgctcggcgcgc 32
||||| ||||||| ||||||| ||||||| |||||
Db 352 CGCGGCGCGCGCTCGCGCGCGCGCTCGC 383

Search completed: August 14, 2002, 21:48:35
Job time: 13533 sec

DR P-PSDB: AAM41370.
XX Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT number of CAG repeat units
XX
PS Claim 1; Page 10; 23pp; Japanese.
XX
CC This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxis type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxis type II.
XX
SQ Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

Query Match 98.8%; Score 31.6; DB 19; Length 355;
Best Local Similarity 96.9%; Pred. No. 0.15;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgccaaccgcgcctcccgctcgccgcgcgc 32
|||||
Db 219 cgccaaccgcgcctcccgctcgccgcgcgcy 250

RESULT 2
AAV17229
ID AAV17229 standard; DNA; 623 BP.
XX
AC AAV17229;
XX
DT 29-JUN-1998 (first entry)
XX
DE SCA2 gene fragment.
XX
KW SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 341..583
FT /*tag= a
FT /note= "SCA2 protein fragment, no stop codon given"
FT
XX
PN WO9803679-A1.
XX
PD 29-JAN-1998.
XX
PF 18-JUL-1996; 96MO-JP01999.
XX
PR 18-JUL-1996; 96MO-JP01999.
XX
PA (SRLS-) SRL INC.
XX
PI Sanpei K, Tsuji S;
XX
DR WPI; 1998-120796/11.
DR P-PSDB: AAM41372.
XX
PT Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT number of CAG repeat units
XX
PS Example 1; Page 11-12; 23pp; Japanese.
XX
CC This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxis type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxis type II.
XX
SQ Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 98.8%; Score 31.6; DB 19; Length 623;
Best Local Similarity 96.9%; Pred. No. 0.14;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgccaaccgcgcctcccgctcgccgcgcgc 32
|||||
Db 219 cgccaaccgcgcctcccgctcgccgcgcgcy 250

RESULT 3
AAV06551
ID AAV06551 standard; DNA; 516 BP.
XX
AC AAV06551;
XX
DT 06-JUL-1998 (first entry)
XX
DE SCA2 gene fragment including CAG repeat region.
XX
KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
KW diagnosis; olivoponto-cerebellar atrophy; ss; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT primer_bind /*tag= a
FT complement (241..257)
FT /*tag= a
FT /note= "primer SCA2-A binding site"
FT 349..366
FT /*tag= b
FT /note= "primer SCA2-B binding site"
FT 499..500
FT exon /*tag= c
FT /note= "predicted splice site"
FT 267..332
FT /*tag= d
FT /note= "CAG repeat region"
FT 267..269
FT /*tag= e
FT /note= "CAG repeat"
FT 270..272
FT /*tag= f
FT /note= "CAG repeat"
FT 273..275
FT /*tag= g
FT /note= "CAG repeat"
FT 276..278
FT /*tag= h
FT /note= "CAG repeat"
FT 279..281
FT /*tag= i
FT /note= "CAG repeat"
FT 282..284
FT /*tag= j
FT /note= "CAG repeat"
FT 285..287
FT /*tag= k
FT /note= "CAG repeat"
FT 291..293
FT /*tag= l
FT /note= "CAG repeat"
FT 294..296
FT /*tag= m
FT /note= "CAG repeat"
FT 297..299
FT /*tag= n
FT /note= "CAG repeat"
FT 300..302
FT /*tag= o
FT /note= "CAG repeat"
FT 306..308
FT repeat_unit

FT		/tag= p
FT	repeat_unit	/note= "CAG repeat"
FT		309..311
FT		/tag= q
FT	repeat_unit	/note= "CAG repeat"
FT		312..314
FT		/tag= r
FT	repeat_unit	/note= "CAG repeat"
FT		315..317
FT		/tag= s
FT	repeat_unit	/note= "CAG repeat"
FT		318..320
FT		/tag= t
FT	repeat_unit	/note= "CAG repeat"
FT		321..323
FT		/tag= u
FT	repeat_unit	/note= "CAG repeat"
FT		324..326
FT		/tag= v
FT	repeat_unit	/note= "CAG repeat"
FT		327..329
FT		/tag= w
FT	repeat_unit	/note= "CAG repeat"
FT		330..332
FT		/tag= x
FT	repeat_unit	/note= "CAG repeat"
PN		W09742314-A1.
PD		13-NOV-1997.
PE		08-MAY-1997; 97WO-US07725.
PX		08-OCT-1996; 96US-0727084.
PX		PR 08-MAY-1996; 96US-0017388.
PR		19-JUL-1996; 96US-0022207.
PA	(CEDA-) CEDARS SINAI MEDICAL CENT.	
PI	Pulst S;	
XX		WPT; 1998-086523/08.
XX	Nucleic acids encoding human and mouse ataxin 2 - a product of the	
XX	sphiocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of	
XX	ataxia type 2	
PS	Example 2; Page 51-52; 98pp; English.	
CC	This genomic DNA in plasmid p165122B includes a CAG repeat region	
CC	from the novel human SCA2 gene (see AAV06552). It was identified	
CC	following the construction of a bacterial artificial chromosome	
CC	contig and a pl artificial chromosome of the sphiocerebellar	
CC	ataxia 2 (SCA2) gene region and the identification of the SCA2	
CC	gene from this contiguous map unit using a technique that screens	
CC	for the presence of DNA trinucleotide repeats. The SCA2 locus is	
CC	at 12q24.1. Ataxia type 2 can be diagnosed by detecting a genomic	
CC	or transcribed mRNA sequence in an individual having an expanded	
CC	CAG repeat at a location corresponding to the CAG repeat region of	
CC	the SCA2 gene. The presence of at least 13 CAG repeats above the	
CC	normal level (22, occasionally 23, repeats) is indicative of SCA2.	
CC	primers (see AAT99640-41) amplifying at least this region are used	
CC	for diagnosis. Also claimed are full-length ataxin-2 cDNAs for	
CC	human and mouse (see AAU06552-53), kits for detecting mutations abt	
CC	the SCA2 locus, antisense oligonucleotides, and transgenic animals	
CC	useful for studying the physiological roles of SCA2 polypeptide	
CC	(ataxin-2, see AAW33807-08) and its effect upon behaviour.	
SQ	Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;	

Query Match	96.98;	Score 31;	DB 19;	Length 516;
Best Local Similarity	100.08;	Pred. No. 0.23;		

	Matches	31; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy	1	cgcacaaccgcgctccgccgtcggagccc	31						
Dd	130	cgcacaaccgcgctccgccgtcggagccc	160						

RESULT	4
AAT78912	ID AAT78912 standard; CDNA: 4200 BP.
XX AC	
XX XX	AAT78912;
DT DE	09-FEB-1998 (first entry)
XX DE	Spinocerebellar ataxia gene SCA2.
XX KM	Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;
XX KM	repeat region; affinity; YARA binding protein; Kennedy disease;
KM KM	transcription initiation factor; lymphoblastic cell line; schizophrenia;
KM KM	Huntington's disease; dominant autosomal spinocerebellar ataxia;
KM KM	X-linked spino-bulbar muscular atrophy; familial spastic paraplegia;
KM KM	dentatorubral-pallidoluysial atrophy; bipolar affective disorder;
XX KM	manic depressive psychosis; ss.
OS OS	Homo sapiens.
XX FH	
FH FT	Location/Qualifiers
FT FT	3..2747
FT FT	/tag= a
FT FT	/product= SCA2 protein
FT FT	/note= "this CDS contains a putative translational start-
FT FT	codon for the SCA2 protein at positions 243-245"
FT FT	2594..3640
FT FT	/tag= b
FT FT	/note= "this second open reading frame may be derived
FT FT	by a frameshift or by alternative splicing"
FT FT	3..242
FT FT	/tag= c
FT FT	/note= "putative open reading frame which is in frame
FT FT	with the putative translational start site of
FT FT	the SCA2 open reading frame"
FT FT	239..245
FT FT	/tag= d
FT FT	/note= "putative Kozak consensus signal"
FT FT	258..323
FT FT	/tag= e
FT FT	/note= "encodes polyglutamine repeat region; contains
FT FT	repeats of CAG with 2 CAA codons interspersed"
FT FT	258..260
FT FT	/tag= f
FT FT	/note= "CAG repeats"
FT FT	1..3986
FT FT	/tag= g
FT FT	/note= "sequence contained in DANI clone"
FT FT	3987..4200
FT FT	/tag= h
FT FT	/note= "derived from the EST's AAH92640, AAN90240 and
FT FT	AAI13574 from dbEST database"
FT FT	4023..4029
FT FT	/tag= i
FT FT	/note= "region which differs in length between the
FT FT	sequences of the EST clones AAH92640, AAN90240
FT FT	and AAI13574"
XX PN	WO9717445-A1.
XX PD	15-MAY-1997.
XX PF	08-NOV-1996:
XX PR	96MO-FR01773.
PR	10-NOV-1995; 95FR-0013576.

PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Lutz Y, Mandel J, Tora L, Trottier Y;
 XX
 DR WPI: 1997-281034/25.
 DR P-PSDB: AAM24800, AAM24801.
 XX
 PT Antibody 1C2 used for treating or preventing neuro-degenerative
 PT diseases - associated with proteins containing long poly:glutamine
 PT repeats, e.g. Huntington's disease
 PS
 PS Claim 21: Page 45-47: 69pp: French.
 XX
 CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
 CC treatment of neurodegenerative diseases associated with the presence
 CC of polyglutamine repeat regions. This Mab is already known for its
 CC affinity to the TATA binding protein (TBP) transcription initiation
 CC factor, especially at the amino acid sequence LEEQKROQKQKQ found at
 CC the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
 CC for polyglutamine repeats with a proportional affinity to the number
 CC of glutamine repeats. This affinity has been used to identify genes
 CC encoding proteins containing long polyglutamine repeats which are
 CC implicated in neurodegenerative diseases. A screen of an expression
 CC library, generated from a lymphoblastic cell line from a patient
 CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
 CC new sequences (AA178906-178911) encoding polyglutamine repeats. Mab 1C2
 CC also isolated the complete SCA2 gene in clone DANI (sequence presented
 CC here). The sequence appears to contain 2 open reading frames (ORF) the
 CC second of which may be generated by an frameshift slippage or by an
 CC alternative splicing event. The first ORF also encodes a 22 amino acid
 CC polyglutamine repeat region near the N-terminus with 1-3 CAA repeats
 CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
 CC interspersed whereas the mutant sequence from patients with SCA
 CC contains at least 30, preferably 37-50 CAG repeats.
 CC Mab 1C2, active fragment of it or nucleic acids encoding it are
 CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
 CC X-linked spinobulbar muscular atrophy (Kennedy disease),
 CC dentatorubral pallidolusar atrophy, dominant autosomal spinocerebellar
 CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
 CC depressive psychoses and schizophrenia.
 CC
 SQ Sequence 4200 BP: 1152 A; 1200 C; 913 G; 935 T; 0 other:

Query Match 96.9%; Score 31; DB 18; Length 4200;
 Best Local Similarity 100.0%; Pred.No. 0.17;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccaaccgcgcctcccgctgcgcgcg 31
 ||||||||||||||||||||||||||||
 DB 121 cgccaaccgcgcctcccgctgcgcgcg 151

RESULT 5
 AAV30270
 ID AAV30270 standard; DNA: 4367 BP.
 AC AAV30270:
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
 XX
 KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
 KW CAG repeat; neurodegenerative disease; ds.
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key location/Qualifiers
 FT CDS 49..3990
 FT /*tag= a
 FT /product= "Spinocerebellar ataxia type 2 associated

FT repeat_region 544..612 protein"
 FT /*tag= b
 FT /note= "normal CAG repeat region; this is increased in
 FT patients with SCA2"
 FT repeat_unit 544..546
 FT /*tag= c
 XX
 XX MO9818920-A1.
 XX
 XX 07-MAY-1998.
 PD
 XX
 XX 30-OCT-1997; 97MO-JP03946.
 PE
 XX
 XX 30-OCT-1996; 96JP-0304059.
 PR
 XX
 XX (SRLS-) SRL INC.
 PA
 XX
 PI Sample K, Tsuji S;
 XX
 DR WPI: 1998-272215/24.
 DR P-PSDB: AAM60213.
 XX
 PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
 PT - contain increased number of CAG repeat region compared to normal
 PT gene
 PS
 PS Claim 1: Pages 13-22: 38pp: Japanese.
 XX
 XX This represents the sequence of a gene causative of spinocerebellar
 CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
 CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
 CC expression product produces a polyglutamine sequence from Gln-166 to
 CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
 CC patients this number is increased to 35-100. Peptides encoded by nucleic
 CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
 CC gene, antibodies recognising the peptides and antisense nucleic acids
 CC hybridising with the nucleic acid fragments can be used for the
 CC investigation and diagnosis of SCA2. They can also be used for the
 CC treatment of SCA2 by antisense therapy or gene therapy.
 CC
 SQ Sequence 4367 BP: 1124 A; 1328 C; 991 G; 924 T; 0 other:

Query Match 96.9%; Score 31; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred.No. 0.17;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccaaccgcgcctcccgctgcgcgcg 31
 ||||||||||||||||||||||||||||
 DB 407 cgccaaccgcgcctcccgctgcgcgcg 437

RESULT 6
 AAV06552
 ID AAV06552 standard; cDNA: 4481 BP.
 AC AAV06552:
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human SCA2 cDNA including CAG repeat region.
 XX
 KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KW diagnosis; olivo-ponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key location/Qualifiers
 FT CDS 164..4101
 FT /*tag= a
 FT primer_bind complement (631..648)
 FT /*tag= b

```

FT /note= "primer SCA2-A binding site"
FT primer_bind 740..757
FT /tag= C
FT /note= "primer SCA2-B binding site"
FT primer_bind 1070..1091
FT /tag= d
FT /note= "primer SCA2-14B binding site"
FT exon 899..900
FT /tag= e
FT /note= "predicted splice site"
FT repeat_region 658..723
FT /tag= f
FT /note= "CAG repeat region"
FT repeat_unit 658..660
FT /tag= g
FT /note= "CAG repeat"
FT repeat_unit 661..663
FT /tag= h
FT /note= "CAG repeat"
FT repeat_unit 664..666
FT /tag= i
FT /note= "CAG repeat"
FT repeat_unit 667..669
FT /tag= j
FT /note= "CAG repeat"
FT repeat_unit 670..672
FT /tag= k
FT /note= "CAG repeat"
FT repeat_unit 673..675
FT /tag= l
FT /note= "CAG repeat"
FT repeat_unit 676..678
FT /tag= m
FT /note= "CAG repeat"
FT repeat_unit 679..681
FT /tag= n
FT /note= "CAG repeat"
FT repeat_unit 685..687
FT /tag= o
FT /note= "CAG repeat"
FT repeat_unit 688..690
FT /tag= p
FT /note= "CAG repeat"
FT repeat_unit 691..693
FT /tag= q
FT /note= "CAG repeat"
FT repeat_unit 694..696
FT /tag= r
FT /note= "CAG repeat"
FT repeat_unit 700..702
FT /tag= s
FT /note= "CAG repeat"
FT repeat_unit 703..705
FT /tag= t
FT /note= "CAG repeat"
FT repeat_unit 706..708
FT /tag= u
FT /note= "CAG repeat"
FT repeat_unit 709..711
FT /tag= v
FT /note= "CAG repeat"
FT repeat_unit 712..714
FT /tag= w
FT /note= "CAG repeat"
FT repeat_unit 715..717
FT /tag= x
FT /note= "CAG repeat"
FT repeat_unit 718..720
FT /tag= y
FT /note= "CAG repeat"
FT repeat_unit 721..723
FT /tag= z
FT /note= "CAG repeat"

```

```

XX MO9742314-A1.
XX
XX PD 13-NOV-1997.
XX
XX PF 08-MAY-1997; 97WO-US07725.
XX
XX PR 08-OCT-1996; 96US-0727084.
XX PR 08-MAY-1996; 96US-0017388.
XX PR 19-JUL-1996; 96US-0022207.
XX
XX PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
XX PI Pulst S;
XX
XX DR WPI; 1998-086523/08.
XX DR P-PSDB; AAW33807.
XX
XX PT Nucleic acids encoding human and mouse ataxin 2 - a product of the
XX PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
XX PT ataxia type 2
XX
XX PS Clalm 6; Page 52-58; 98pp; English.
XX
XX CC This cDNA sequence corresponds to a novel SCA2 gene encoding a human
XX CC spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
XX CC (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
XX CC human frontal cortex cDNA library in lambda ZapII were screened
XX CC with probes obtained by PCR amplification of plasmid AAW61228 (see
XX CC AAW06551). PCR products were used to screen the human adult frontal
XX CC cortex library and 5' clones were obtained by RT-PCR of placental
XX CC mRNAs. Overlapping clones was used to generate the composite 4481
XX CC bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
XX CC or transcribed mRNA sequence in an individual having an expanded
XX CC CAG repeat at a location corresponding to the CAG repeat region of
XX CC the SCA2 gene. The presence of at least 13 CAG repeats above the
XX CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
XX CC Primers (see AAW9640-41) amplifying at least this region are used
XX CC for diagnosis. Also claimed are kits for detecting mutations at
XX CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
XX CC useful for studying the physiological roles of ataxin-2 and its
XX CC effect upon behaviour.
XX
XX SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 96.9%; Score 31; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. NO. 0.17;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgccaaccgcgcgtcccgctcgccgcg 31
Db 521 cgccaaccgcgcgtcccgctcgccgcg 551

RESULT 7
AAZ23428
ID AAZ23428 standard; DNA: 4481 BP.
XX
XX AC AAZ23428;
XX
XX DT 19-JAN-2000 (first entry)
XX
XX DE Human SCA2 DNA.
XX
XX KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
XX KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
XX KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
XX KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
XX KW dentatorubropallidolysian atrophy; cell proliferation; cell survival;
XX KW neoplastic; malignant; autoimmune; fibrotic; ss.
XX
XX OS Homo sapiens.

```

```

XX Key Location/Qualifiers
FH CDS 163..4101
FT /tag= a
FT /product= "SCA2"
XX
XX MO9945944-A1.
XX
XX 16-SEP-1999.
XX
XX 11-MAR-1999: 99WO-US05250.
XX
XX 12-MAR-1998: 98US-0041886.
XX
XX (BURN-) BURNHAM INST.
XX
XX Bredesen DE, Rabizadeh S;
XX
XX WPI: 1999-561617/47.
XX
XX P-PSDB: AAY33495.
XX
XX New proapoptotic dependence peptides, used to develop products for
XX treating, e.g. Alzheimer's disease -
XX
XX PS Disclosure: Page 130-135; 1999p: English.
XX
XX This invention describes novel pure proapoptotic dependence peptides
XX which comprise a sequence of an active dependence domain selected from
XX dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
XX huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
XX SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
XX of inducing cell death and can be used to develop products to mediate or
XX inhibit apoptosis. The methods can be used for reducing the severity of
XX a proapoptotic dependence domain mediated pathological conditions e.g.
XX Huntington's disease, Alzheimer's disease, Kennedy's disease,
XX spinocerebellar ataxias, dentatorubropallidolysian atrophy,
XX Machado-Joseph disease, stroke or head trauma. They can also be used for
XX reducing the severity of a pathological condition mediated by upregulated
XX cell proliferation or cell survival e.g. neoplastic, malignant,
XX autoimmune or fibrotic conditions. This sequence encodes the human
XX SCA2 polypeptide described in the method of the invention.
XX
XX SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other:
XX
XX Query Match 96.9%; Score 31; DB 20; Length 4481;
XX Best Local Similarity 100.0%; Pred. No. 0.17;
XX Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 cgccaaccgcgcctcccgctcgagccgcg 31
XX ||||||||||||||||||||||||||||
XX Db 521 cgccaaccgcgcctcccgctcgagccgcg 551
XX
XX RESULT 8
XX AAS46300/c
XX ID AAS46300 standard; DNA; 6862 BP.
XX
XX AC AAS46300;
XX
XX DT 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #22.
XX
XX DE Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
XX KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX OS Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.

```

```

XX 15-MAR-2001; 2001WO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX
XX 06-APR-2000; 2000DE-1019058.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX analysing diseases associated with cytosine methylation state e.g.
XX cancer -
XX
XX PS Claim 1; SEQ ID No 22; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (SS) and sequences complementary to (SS). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases. By analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes. Sequences with even numbered Seq ID numbers are the
XX complementary sequence of the corresponding odd numbered sequence (e.g.
XX CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
XX is missing).
XX
XX CC Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other:
XX
XX Query Match 68.8%; Score 22; DB 22; Length 6862;
XX Best Local Similarity 83.3%; Pred. No. 96;
XX Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 1 cgccaaccgcgcctcccgctcgagccgcg 30
XX ||||||||||||||||||||||||
XX Db 278 CGACCAACCGCGCGCGCGCGCACGCGCCG 249
XX
XX RESULT 9
XX ABL32223/c
XX ID ABL32223 standard; DNA; 6862 BP.
XX
XX AC ABL32223;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 196.
XX
XX DE Human: immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antihaemic; cytostatic; nootropic;

```

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 OS Homo sapiens.
 PN WO200200928-A2.
 XX
 XX 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1; SEQ ID NO 196; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SO Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other;
 Query Match 68.8%; Score 22; DB 24; Length 6862;
 Best Local Similarity 83.3%; Pred. No. 96;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 cgcgaaccgcgcctcccgctggcgccc 30
 DB 278 CGACAAACCGCGCGCGCGACGCGCC 249
 RESULT 10
 ID AAS61082/C
 XX AAS61082 standard; DNA: 6862 BP.
 AC
 XX
 DT 29-JAN-2002 (first entry)
 DE Human gene regulation-associated gene oligonucleotide #37.
 XX
 KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX

PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03968.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-017470/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated
 PT with gene regulation, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
 PT disease -
 PS Claim 1; SEQ ID NO 38; 26pp; English.
 XX
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The
 CC chemical pretreatment converts cytosine bases unmethylated at the
 CC 5-position to uracil or another base with hybridisation behaviour
 CC dissimilar to cytosine, to enable analysis of cytosine methylations.
 CC The DNA sequences, oligomers (or sets/arrays) and method are
 CC useful in the diagnosis of diseases (or predisposition to diseases)
 CC associated with gene regulation and in therapy of such diseases, by
 CC enabling analysis of the cytosine methylation patterns of such genes,
 CC kits are provided. They are especially useful in diagnosis
 CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
 CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
 CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
 CC preeclampsia, graft versus-host disease. The present sequence is a
 CC sequence included in the sequence data for this specification and is
 CC associated with the human gene regulation-associated genes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp://ipo.int/pub/published_pcl_sequences
 XX
 SO Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other;
 Query Match 68.8%; Score 22; DB 24; Length 6862;
 Best Local Similarity 83.3%; Pred. No. 96;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 cgcgaaccgcgcctcccgctggcgccc 30
 DB 278 CGACAAACCGCGCGCGCGACGCGCC 249
 RESULT 11
 ID AAF67698/C
 XX AAF67698 standard; DNA: 98 BP.
 AC
 XX
 DT 12-APR-2001 (first entry)
 DE Insulator plasmid enhancer blocking sequence ApB SEQ ID NO: 56.
 XX
 KW Chicken; human; insulator; enhancer; DNA binding protein;
 KW gene expression; gene therapy; insulin-like growth factor-2; Igf2;
 KW knockout mouse; ds.
 XX
 OS Unidentified.
 XX
 PN WO200102553-A2.
 XX

CC specifically binds to and cleaves transcribed DNA, which reduces or
CC inhibits synthesis of the protein coding DNA; (3) an expression vector,
CC containing DNA as above, or which encodes antisense RNA or a ribozyme;
CC (4) a host cell transformed with a vector as in (3); (5) a protein,
CC encoded by DNA as above; (6) a method to produce the protein of (5),
CC comprising culturing the cell of (4) and isolating the protein from the
CC cell or the culture medium; (7) an antibody targeted against the protein
CC of (5); (8) a diagnostic method to detect disturbed expression of the
CC protein of (5) or to detect altered forms of the protein by contacting a
CC sample with a DNA sequence or antibody and determining direct or indirect
CC contact, and comparing the expression of the protein with a healthy
CC patient; (9) a diagnostic kit to perform the method of (8); (10) a
CC non-human transgenic animal, where the naturally occurring T gene has an
CC altered gene structure or sequence; and (11) a method to produce a
CC non-human animal as in (10). The DNA, derived from the T gene encodes a T
CC protein (TP) which is involved in development of the central nervous
CC system. Antisense sequences, ribozymes and antibodies are useful for
CC treatment of disorders of the CNS including schizophrenia, autism, manic
CC depression and mental retardation. This sequence encodes a fragment of
CC the human T protein described in the method of the invention.
CC
XX
SQ Sequence 3682 BP; 869 A; 888 C; 933 G; 992 T; 0 other;

Query Match 65.6%; Score 21; DB 21; Length 3682;
Best Local Similarity 82.8%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 cgcgaaccgcgcctccgcgcgcgcgcgcgc 29
||||| ||||||| ||||||| |||||||
DB 2263 CGCCGGCGCGCGCTCCCGCGCGCGCGC 2235

RESULT 14
AAC10674/c
ID AAC10674 standard; cDNA; 146 BP.

AAC10674;

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 14749.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 1; SEQ ID 14749; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from
cDNAs encoding secreted proteins. No ORF has yet been conclusively
identified within the present sequence. The 5' ESTs were prepared from
total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
CC
XX
SQ Sequence 146 BP; 17 A; 51 C; 59 G; 19 T; 0 other;

Query Match 65.0%; Score 20.8; DB 21; Length 146;
Best Local Similarity 78.1%; Pred. No. 3.8e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 cgcgaaccgcgcctccgcgcgcgcgcgcgc 32
||||| ||||||| ||||||| |||||||
DB 42 CGCCATCTCTCTCTGCGACCGCGCTCCGC 11

RESULT 15

AAH07599
ID AAH07599 standard; cDNA; 232 BP.

AAH07599;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:4434.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

Claim 1; SEQ ID 4434; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of

the 5'- and sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95683 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 232 BP; 20 A; 102 C; 79 G; 28 T; 3 other;

Query Match	65.0%	Score 20.8	DB 22	Length 232
Best Local Similarity	78.1%	Pred No. 3.5e+02		
Matches 25	Conservative 0	Mismatches 7	Indels 0	Gaps 0

OY	1	cgcacaccgcgccctcccgcgtcggcgcgcgc	32
Db	103	cgccacactgagctctccgcgcgcgcgcgcc	134

Search completed: August 14, 2002, 22:06:40
Job time: 11695 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:55:21 ; Search time 203.42 seconds

(without alignments)
38.641 Million cell updates/sec

Title: US-09-707-919-6

Sequence: 32

Sequence: 1 cgcacaccgcgcctcccgctcgccgcgcgcgc 32

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.6	98.8	355	4	US-09-043-303-1
2	31.6	98.8	623	4	US-09-043-303-5
3	31.6	98.9	4481	4	US-09-041-886-18
4	20.8	65.0	3001	4	US-09-387-212-9
5	20.8	65.0	4171	1	US-08-308-881-5
6	20.8	65.0	4171	2	US-09-058-263-5
7	20.8	65.0	4171	2	US-09-059-099-5
8	20.8	65.0	4171	3	US-09-058-264-5
9	20.8	65.0	4171	5	PCT-US95-06530-5
10	19.6	61.3	1729	4	US-09-045-973-6
11	19.6	61.3	2647	5	PCT-US93-06251-77
12	19.4	60.6	17138	4	US-09-813-819-3
13	19.4	60.6	17138	4	US-09-920-048-3
14	19.2	60.0	1479	1	US-08-644-221-31
15	19.2	60.0	2538	3	US-08-899-437-1
16	19.2	60.0	2538	4	US-09-126-121-1
17	19.2	60.0	2574	2	US-08-677-734A-8
18	19.2	60.0	3663	4	US-09-499-884-11
19	19.2	59.4	50937	4	US-09-428-517-1
20	18.8	58.8	220	4	US-09-094-207A-11
21	18.8	58.8	3013	2	US-09-096-982-6
22	18.8	58.8	3013	2	US-08-653-650A-6
23	18.8	58.8	3804	2	US-08-483-488-5
24	18.8	58.8	9595	3	US-09-014-416-4
25	18.6	58.1	1335	4	US-09-045-973-2
26	18.6	58.1	1491	4	US-09-082-092-9
27	18.6	58.1	1817	4	US-09-288-292A-45

28	18.6	58.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
29	18.4	57.5	220	4	US-09-060-756-593	Sequence 593, App
30	18.4	57.5	1018	1	US-08-444-083-7	Sequence 7, Appl1
31	18.4	57.5	1018	1	US-08-286-304-7	Sequence 7, Appl1
32	18.4	57.5	1018	1	US-08-442-745-7	Sequence 7, Appl1
33	18.4	57.5	1018	1	US-08-443-129-7	Sequence 7, Appl1
34	18.4	57.5	1018	1	US-08-443-952-7	Sequence 7, Appl1
35	18.4	57.5	1018	1	US-08-443-130-7	Sequence 7, Appl1
36	18.4	57.5	1018	3	US-08-898-911-7	Sequence 7, Appl1
37	18.4	57.5	1018	5	PCT-US95-04467-7	Sequence 7, Appl1
38	18.4	57.5	1157	1	US-07-709-949-1	Sequence 7, Appl1
39	18.4	57.5	1529	3	US-08-858-876A-3	Sequence 3, Appl1
40	18.4	57.5	1329	4	US-09-472-880-3	Sequence 3, Appl1
41	18.4	57.5	44377	2	US-08-804-227C-7	Sequence 7, Appl1
42	18.4	57.5	44377	2	US-08-804-198-1	Sequence 7, Appl1
43	18.4	57.5	50341	1	US-08-247-901C-1	Sequence 1, Appl1
44	18.4	57.5	50341	2	US-09-075-904-1	Sequence 1, Appl1
45	18.4	57.5	52297	4	US-09-426-436-1	Sequence 1, Appl1

ALIGNMENTS

```
RESULT 1
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
US-09-043-303-1

Query Match          98.8%: Score 31.6; DB 4; Length 355;
Best Local Similarity 96.9%: Pred. No. 0.022;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcacaccgcgcctcccgctcgccgcgcgc 32
|||||
DB 219 cgcacaccgcgcctcccgctcgccgcgcgy 250

RESULT 2
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
```

```

: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 5
: LENGTH: 623
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (341)..(583)
: FEATURE:
: OTHER INFORMATION: Tsp-2
US-09-043-303-5

Query Match          98.8%: Score 31.6; DB 4; Length 623;
Best Local Similarity 96.9%: Pred. No. 0.021;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccaaccgcgcctcccgctcgagccgcg 32
    ||||||||||||||||||||||||||||
Db 219 cgccaaccgcgcctcccgctcgagccgcg 250

RESULT 3
US-09-041-886-18
: Sequence 18, Application US/09041886
: Patent No. 6235872
: GENERAL INFORMATION:
: APPLICANT: Bredesen, Dale E.
: APPLICANT: Rabizadeh, Sharroo
: TITLE OF INVENTION: Proapoptotic Peptides, Dependence
: TITLE OF INVENTION: Polypeptides and Methods of Use
: NUMBER OF SEQUENCES: 72
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/041.886
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 2626
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4481 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 163..4099
US-09-041-886-18

Query Match          96.9%: Score 31; DB 4; Length 4481;
Best Local Similarity 100.0%: Pred. No. 0.027;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgccaaccgcgcctcccgctcgagccgcg 31
```

```

Db 521 CGCCAACCGCGCCTCCCGCTCGAGCCCGC 551
    ||||||||||||||||||||||||||||
RESULT 4
US-09-387-212-9
: Sequence 9, Application US/09387212A
: Patent No. 6309849
: GENERAL INFORMATION:
: APPLICANT: ROBISON, KEITH E.
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
: TITLE OF INVENTION: PHOSPHATASE HOMOLOGUES AND USES THEREFOR
: FILE REFERENCE: MNI-090
: CURRENT APPLICATION NUMBER: US/09/387,212A
: CURRENT FILING DATE: 1999-08-31
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 9
: LENGTH: 3001
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-387-212-9

Query Match          65.0%: Score 20.8; DB 4; Length 3001;
Best Local Similarity 78.1%: Pred. No. 42;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 cgccaaccgcgcctcccgctcgagccgcg 32
    |||| |||| || |||| || |||| ||
Db 33 cgctcccgccgcgcgcgcgcgcgcgcgcg 64

RESULT 5
US-08-308-881-5
: Sequence 5, Application US/08308881
: Patent No. 5783672
: GENERAL INFORMATION:
: APPLICANT: Mosley, Bruce
: APPLICANT: Cosman, David J.
: TITLE OF INVENTION: Receptor for Oncostatin M
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Apple 7.1
: SOFTWARE: Microsoft Word, Version 5.1a
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/308,881
: FILING DATE: 12-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/249,553
: FILING DATE: 26-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Seese, Kathryn A.
: REGISTRATION NUMBER: 32,172
: REFERENCE/DOCKET NUMBER: 2614-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 587-0430
: TELEFAX: (206) 233-0644
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4171 base pairs
: TYPE: nucleic acid
```

```

1 FILLING DATE:
2 CLASSIFICATION:
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US/08/308,881
5 FILLING DATE: 12-SEP-1994
6 APPLICATION NUMBER: US 08/249,553
7 FILLING DATE: 26-MAY-1994
8 ATTORNEY/AGENT INFORMATION:
9 NAME: Seese, Kathryn A.
10 REGISTRATION NUMBER: 32,172
11 REFERENCE/DOCKET NUMBER: 2614-A
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (206) 587-0430
14 TELEFAX: (206) 233-0644
15 TELEX: 756822
16 INFORMATION FOR SEQ ID NO: 5:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 4171 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: cDNA to mRNA
23 HYPOTHEICAL: NO
24 ANTI-SENSE: NO
25 IMMEDIATE SOURCE:

```

CLONE: huosm-Ra
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 368..448
FEATURE:
NAME/KEY: CDS
LOCATION: 368..3307
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 449..3304
US-09-059-099-5

Query Match 65.0%; Score 20.8; DB 2; Length 4171;
Best Local Similarity 78.1%; Pred. No. 41;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 cgcaccgcgcctcccgctcgagcgccgc 32
DB 102 CCCGACCCGCCCTCCCGCTGCTGCGGC 133

RESULT 8
US-09-058-264-5
Sequence 5, Application US/09058264
Patent No. 6010886
GENERAL INFORMATION:
APPLICANT: Mosley, Bruce
APPLICANT: Cosman, David J.
TITLE OF INVENTION: Receptor for Oncostatin M
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,264
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/308,881
FILING DATE: 12-SEP-1994
APPLICATION NUMBER: US 08/249,553
FILING DATE: 26-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2614-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: huosm-Ra
FEATURE:
NAME/KEY: sig_peptide

LOCATION: 368..448
FEATURE:
NAME/KEY: CDS
LOCATION: 368..3307
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 449..3304
US-09-058-264-5

Query Match 65.0%; Score 20.8; DB 3; Length 4171;
Best Local Similarity 78.1%; Pred. No. 41;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 cgcaccgcgcctcccgctcgagcgccgc 32
DB 102 CCCGACCCGCCCTCCCGCTGCTGCGGC 133

RESULT 9
PCT-US95-06530-5
Sequence 5, Application PC/TUS9506530
GENERAL INFORMATION:
APPLICANT: Mosley, Bruce
APPLICANT: Cosman, David J.
TITLE OF INVENTION: Receptor for Oncostatin M
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06530
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,881
FILING DATE: 09-SEP-1994
APPLICATION NUMBER: US 08/249,553
FILING DATE: 26-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Anderson, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2614-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: huosm-R'
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 368..448
FEATURE:
NAME/KEY: CDS
LOCATION: 368..3307

```

; FEATURE: mat_peptide
; NAME/KEY: 449..3304
; LOCATION:
;
PCT-US95-06530-5

```

Query Match	65.0%;	Score 20.8;	DB 5;	Length 4171;
Best Local Similarity	78.1%;	Pred. No. 41;		
Matches 25; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;

Oy	1	cgccaacccgcgcctccccgctcgcgcgcgc	32
Db	102	ccccgacccgcccgtcccccgtctgcgcgc	133

RESULT 10
US-09-045-973-6
; Sequence 6, Application US/09045973

APPLICANT: Lal, Preeti
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah
TITLE OF INVENTION: PROTEIN PHOSPHATASE RELATED MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSD for Windows Vers1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/045.973
 FILING DATE: Filed Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0491 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 TELEX:

```

: INFORMATION FOR SEQ ID NO: 6 :
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1729 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: BRSTNOT16
: CLONE: 3041794
:
US-09-045-973-6

```

```

Oy      4 caaccgcgcctcccgctcgcgcc 29
          ||||| | | | | | | | |
Db      285 CAACCCGGCGCGCGCGCGCGCGCC 310

```

RESULT 11
PCT-US93-06251-77
; Sequence 77, Application PC/TUS9306251

```

1  APPLICANT:  Wickstrom, Eric and Rife, Jason P.
2  TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
3  TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
4  NUMBER OF SEQUENCES:  93

```

ADDRESS: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patentin Release #1.0, Version #1.2
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: PCT/US93/06251
8  FILING DATE: 19930630

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8588
; TELECOMMUNICATION INFORMATION:

```

```

? INFORMATION FOR SEQ ID NO: 77:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 2647 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: double
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
PCT-US93-06251-77

```

Query Match	61.3%	Score 19.6	DB 5	Length 2647
Best Local Similarity	84.6%	Pred. No. 1e+02		
Matches 22; Conservative		0; Mismatches 4;	Indels 0;	Gaps 0

QY	7	ccccgcctccccgcctgcgcgcgc	32
Db	248	ccccgcgcgcacccgcgcgcgcgcgc	273

```

: RESULT 12
: US-09-813-819-3/C
: Sequence 3, Application US/09813819
: Patent No. 6294368
: GENERAL INFORMATION:
: APPLICANT: MERKULOV, Gennady et al
: TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: CL001177
: CURRENT APPLICATION NUMBER: US/09/813,819
: CURRENT FILING DATE: 2001-03-22
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3

```

LOCATION: (1)...(17138)
OTHER INFORMATION: n = A,T,C or G
US-09-813-819-3

Query Match 60.6%; Score 19.4; DB 4; Length 17138;
Best Local Similarity 79.3%; Pred. No. 98;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 caaccgcgcctcccgctcgagccgc 32
| | | | | | | | | | | | | | | | | | | | | |
Db 2786 CCACCGCGCTGCGCTCTCGTCCGCGC 2758

RESULT 13
US-09-920-048-3/C
Sequence 3, Application US/09920048
Patent No. 6344352
GENERAL INFORMATION:
APPLICANT: MERKULOV, Gennady et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
FILE REFERENCE: CL001177DIV
CURRENT APPLICATION NUMBER: US/09/920,048
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/813,819
PRIOR FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 17138
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(17138)
OTHER INFORMATION: n = A,T,C or G
US-09-920-048-3

Query Match 60.6%; Score 19.4; DB 4; Length 17138;
Best Local Similarity 79.3%; Pred. No. 98;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 caaccgcgcctcccgctcgagccgc 32
| | | | | | | | | | | | | | | | | | | | | |
Db 2786 CCACCGCGCTGCGCTCTCGTCCGCGC 2758

RESULT 14
US-08-644-271-31
Sequence 31, Application US/08644271
Patent No. 5814478
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
TITLE OF INVENTION: AND LIGANDS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271

FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Covert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1476
OTHER INFORMATION:
NAME/KEY: Human Agrin
LOCATION: 1...1479
OTHER INFORMATION:
US-08-644-271-31

Query Match 60.0%; Score 19.2; DB 1; Length 1479;
Best Local Similarity 87.5%; Pred. No. 1,4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 acccgagctcccgctcgagcc 29
| | | | | | | | | | | | | | | | | | | | | |
Db 1264 ACCGAGCTCCCTCCGCTGCGGCC 1287

RESULT 15
US-08-899-437-1
Sequence 1, Application US/08899437
Patent No. 6121415
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,437
FILING DATE: 24-Jul-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Delidre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P1084R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

BASE COUNT 240 a 329 c 306 g 219 t 6 others
 ORIGIN

Query Match 100.0%; Score 32; DB 10; Length 1100;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcgaaccgagcctcccgctcgagcgccgc 32
 ||||||||||||||||||||||||||||||||
 Db 142 CGCCAAACCGCGCTCCCGCTCGGCGCCG 173

RESULT 2
 LOCUS AL039573 482 bp mRNA linear EST 29-FEB-2000
 DEFINITION DKFZP434D1311.F1 434 (synonym: htes3) Homo sapiens cDNA clone
 ACCESSION AL039573
 VERSION AL039573
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann
 S. EST (Duesterhoeft, et al.)
 Unpublished (1999)
 Contact: Duesterhoeft A
 MIPS

Am Klopferplatz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the CDNA sequencing
 consortium of the German Genome Project.
 No s1 sequence available.
 This clone (DKFZP434D1311) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 Source
 Location/Qualifiers
 1..482
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZP434D1311"
 /clone_lib="434 (synonym: htes3)"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSPORT1, Site_1: NotI; Site_2: SalI"

BASE COUNT 49 a 218 c 145 g 70 t

Query Match 96.9%; Score 31; DB 9; Length 482;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcgaaccgagcctcccgctcgagcgccgc 31
 ||||||||||||||||||||||||||||
 Db 168 CGCCAAACCGCGCTCCCGCTCGGCGCCG 198

RESULT 3
 LOCUS BE457923/c 364 bp mRNA linear EST 26-JUL-2000
 DEFINITION BE457923
 IMAGE:3326518 3 similar to TR:070305 070305 SPINOCEREBELLAR ATAXIA
 2 HOMOLOG ;, mRNA sequence.

ACCESSION BE457923
 VERSION BE457923.1 GI:9480561

KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 364)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 This clone is available royally-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1070682
 Possible reversed clone: polyT not found.

FEATURES
 Source
 Location/Qualifiers
 1..364
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3326518"
 /clone_lib="Soares_thymus_2NDMT"
 /sex="male"
 /tissue_type="thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer (5'
 TCTTACCAATCTGAACTGGAGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
 3'); double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 51 a 126 c 173 g 14 t

Query Match 75.0%; Score 24; DB 10; Length 364;
 Best Local Similarity 84.4%; Pred. No. 1.2e+03;
 Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 cgcgaaccgagcctcccgctcgagcgccgc 32
 ||||| |||||||| ||||| ||||| |||||
 Db 289 CGCCGCGCGCGCTCCCGCTCGGCGCCG 258

RESULT 4
 LOCUS BF166472/c 673 bp mRNA linear EST 30-OCT-2000
 DEFINITION BF166472
 mRNA sequence.
 ACCESSION BF166472
 VERSION BF166472.1 GI:11046824
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 673)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

[illegible]

FEATURES									
source									
cdna library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIND at: http://image.llnl.gov Plate: L14M6435 row: b column: 22 High quality sequence stop: 380. location/Qualifiers 1..826 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:346053" /clone_1lb="NH_MGC_12" /tissue_type="cervical carcinoma cell line" /lab_host="DH10B" /note="Organ: cervix; Vector: PCMV-SPOrt6; Site1: NotI; Site2: SalI; Cloned unidirectionally. Primer: oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."									
BASE COUNT									
147 a 384 c 179 g 116 t									
ORIGIN									
Query Match 71.9% Score 23; DB 10; Length 826; Best Local Similarity 83.9% Pred. No. 2.2e+03; Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;									
OY	1	CGCCACCCGCGCCTCCCGCTCGAGCGCG 31 							
Db	443	CGCCACCTGCGCTCCCGCGCGCCCG 473							
RESULT 9									
A2186337 891 bp DNA linear GSS 30-AUG-2000									
LOCUS									
DEFINITION									
SP.1006.B1.D09.T7A Strongylocentrotus purpuratus, purple sea urchin , sperm genomic BAC library Strongylocentrotus purpuratus genomic clone plate=1006 Col-17 Row-H, DNA sequence.									
A2186337									
A2186337.1 GI:8369431									
GSS.									
Strongylocentrotus purpuratus.									
Strongylocentrotus purpuratus									
Eukaryote; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 891) Cameron,R.A., Mallaras,G., Rast,J.P., Martinez,P., Blondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray /G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L. A sea urchin genome project: Sequence scan, virtual map, and additional resources Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)									
20402566									
Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1006 row: H column: 17 Seq primer: T7 Class: BAC ends High quality sequence stop: 891. location/Qualifiers 1..891 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="plate=1006 Col-17 Row-H" /clone_1lb="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"									
FEATURES									
source									

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:48:35 ; Search time 2563.92 Seconds

(without alignments)
261.182 Million cell updates/sec

Title: US-09-707-919-7

Perfect score: 32
Sequence: 1 cgcacccgcgcctcccgctcgcgcccgct 32

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	32	100.0	4163	9	HSNDNSCA2	Y08262 H. sapiens m
2	32	100.0	4200	6	A62706	A62706 Sequence 7
3	32	100.0	4481	6	AR153580	AR153580 Sequence
4	32	100.0	4481	6	HSU70323	U70323 Human ataxi
5	31.6	98.8	355	6	AR159544	AR159544 Sequence
6	31.6	98.8	572	6	AR159558	AR159558 Sequence
7	31.6	98.8	623	6	AR159546	AR159546 Sequence
8	31	96.9	264	9	AF330032	AF330032 Papio ham
9	31	96.9	390	9	AF330028	AF330028 Pan trogl
10	31	96.9	231758	2	AC004085	AC004085 Homo sapi
11	29.4	91.9	303	9	AF330031	AF330031 Macaca mu
12	29.4	91.9	322	9	AF330033	AF330033 Macaca ra
13	29.4	91.9	384	9	AF330030	AF330030 Presbytis
14	28.4	88.8	409	9	AF330029	AF330029 Gorilla g
15	23.6	73.8	29924	1	SC5E9	AL446003 Streptomy
16	23.6	73.8	165242	2	AC095560	AC095560 Rattus no
17	23	71.9	4225	10	AF041472	AF041472 Mus muscu
18	23	71.9	169027	2	AC099282	AC099282 Rattus no
19	22.8	71.2	173967	9	AC073343	AC073343 Homo sapi
20	22.6	70.6	1301	14	S75622	S75622 Immediate e
21	22.6	70.6	1301	14	AF352564	AF352564 Pseudorab
22	22.6	70.6	1411	14	SHU20963	U20963 Suid herpes
23	22.4	70.0	15742	1	AF013216	AF013216 Myxococcu
24	22	68.8	2015	10	DB6548	DB6548 Mouse DNA f
25	22	68.8	6862	6	AX251054	AX251054 Sequence
26	22	68.8	6862	6	AX251777	AX251777 Sequence
27	22	68.8	6862	6	AX345125	AX345125 Sequence
28	22	68.8	21541	2	AC095907	AC095907 Rattus no
29	22	68.8	27709	2	LMFLCHR32_27	Continuation (28 o
30	22	68.8	63082	2	AC022653	AC022653 Homo sapi
31	22	68.8	104943	2	AC103066	AC103066 Rattus no
32	22	68.8	118711	2	AC106401	AC106401 Rattus no
33	22	68.8	143670	2	AC096451	AC096451 Rattus no
34	22	68.8	158623	30	AC021815	AC021815 Homo sapi
35	22	68.8	160583	2	AC109348	AC109348 Homo sapi
36	22	68.8	177285	9	AC079115	AC079115 Homo sapi
37	21.8	68.1	1015	3	AF271281	AF271281 Rhinipiceph
38	21.8	68.1	146150	2	AC068333	AC068333 Homo sapi
39	21.6	67.5	28500	1	SC5B8	AL022374 Streptomy
40	21.6	67.5	41782	1	SCG11A	AL133210 Streptomy
41	21.6	67.5	134662	2	AC109786	AC109786 Bos tauru
42	21.6	67.5	141089	2	AC109915	AC109915 Bos tauru
43	21.4	66.9	1658	10	AX068169	AX068169 Sequence
44	21.4	66.9	1658	10	RATPLPA1	M34108 Rat parathy
45	21.4	66.9	1792	1	AY043329	AY043329 Streptomy

ALIGNMENTS

RESULT 1	HSNDNSCA2	4163 bp	mrna	linear	PRI 09-JAN-1997
LOCUS	HSNDNSCA2				
DEFINITION	H.sapiens mRNA for SCA2 protein.				
ACCESSION	Y08262				
VERSION	Y08262.1	GI:1770389			
KEYWORDS	SCA2 gene.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	Imbert,G., Saudou,F., Yvert,G., Devys,D., Trottier,Y., Gardier,J.M., Weber,C., Mandel,J.L., Cancel,G., Abbas,N., Duerr,A., Didierjean,O., Stevanin,G., Agid,Y. and Brice,A.				
TITLE	Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats				
JOURNAL	Nat. Genet. 14 (3), 285-291 (1996)				
MEDLINE	97051922				
REFERENCE	2 (bases 1 to 4163)				
AUTHORS	Imbert,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-SEP-1996) G. Imbert, I.G.B.M.C., Departement Of				

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: OG

Center clone name: RP11-42B1

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329

Consensus quality: 224788 bases at least Q40

Consensus quality: 229074 bases at least Q30

Consensus quality: 230948 bases at least Q20

Estimated insert size: 227237; sum-of-contigs estimation

Estimated insert size: 317311; agarose-ftp estimation

Quality coverage: 6.3x in Q20 bases; agarose-ftp estimation

Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 33241: contig of 33241 bp in length
* 33242 33341: gap of unknown length
* 33342 36391: contig of 23050 bp in length
* 36392 56491: gap of unknown length
* 56492 81323: contig of 24832 bp in length
* 81324 81423: gap of unknown length
* 81424 102538: contig of 21115 bp in length
* 102539 102538: gap of unknown length
* 102639 119710: contig of 17072 bp in length
* 119711 119810: gap of unknown length
* 119811 136913: contig of 17103 bp in length
* 136914 137013: gap of unknown length
* 137014 153285: contig of 16272 bp in length
* 153286 153385: gap of unknown length
* 153386 167987: contig of 14602 bp in length
* 167988 168087: gap of unknown length
* 168088 178731: contig of 10644 bp in length
* 178732 178831: gap of unknown length
* 178832 186741: contig of 7810 bp in length
* 186742 186741: gap of unknown length
* 186742 193215: contig of 6474 bp in length
* 193216 193315: gap of unknown length
* 193316 201310: contig of 7995 bp in length
* 201311 201410: gap of unknown length
* 201411 208647: contig of 7237 bp in length
* 208648 208747: gap of unknown length
* 208748 213802: contig of 5055 bp in length
* 213803 213902: gap of unknown length
* 213903 218049: contig of 4147 bp in length
* 218050 218149: gap of unknown length
* 218150 223316: contig of 5167 bp in length
* 223317 223416: gap of unknown length
* 223417 227389: contig of 3973 bp in length
* 227390 227489: gap of unknown length
* 227490 229032: contig of 1543 bp in length
* 229033 229132: gap of unknown length
* 229133 230651: contig of 1519 bp in length
* 230652 230751: gap of unknown length
* 230752 231758: contig of 1007 bp in length.
Location/Qualifiers
1..231758

```

FEATURES

Source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP11-42B1"

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others

ORIGIN

Query Match 96.9%; Score 31; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcgaacccgcgcctcccgctgcgcgcgcg 31
|||||
DB 89265 CGCCACCGCGCGCTCCGCGTCCGCGCGCCG 89235

RESULT 11

AF330031 303 bp DNA linear PRI 08-NOV-2001

LOCUS Macaca mulatta SCA2 gene, partial sequence.

DEFINITION AF330031

ACCESSION AF330031.1 GI:12382833

VERSION

KEYWORDS

SOURCE

ORGANISM

thesus monkey.

Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

Cercopithecoidea; Macaca.

REFERENCE

AUTHORS

1 (bases 1 to 303)

Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and

Brahmachari, S.K.

CAG repeat instability at SCA2 locus: anchoring CAA interruptions

and linked single nucleotide polymorphisms

Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

REFERENCE

AUTHORS

2 (bases 1 to 303)

Choudhry, S. and Brahmachari, S.K.

Direct Substitution

Submitted (21-DEC-2000) Functional Genomics Unit, Center for

Biochemical Technology, Delhi University Campus, Mall Road, Delhi

110 007, India

Location/Qualifiers

1..303

/organism="Macaca mulatta"

/db_xref="taxon:9544"

<1..>303

/gene="SCA2"

/note="spinocerebellar ataxia 2"

BASE COUNT 32 a 143 c 92 g 36 t

ORIGIN

Query Match 91.9%; Score 29.4; DB 9; Length 303;

Best Local Similarity 96.8%; Pred. No. 28;

Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cgcgaacccgcgcctcccgctgcgcgcgcg 31

|||||

DB 67 CGCCACCGCGCGCTCCGCGTCCGCGCGCCG 97

RESULT 12

AF330033 322 bp DNA linear PRI 08-NOV-2001

LOCUS Macaca radiata SCA2 gene, partial sequence.

DEFINITION AF330033

ACCESSION AF330033.1 GI:12382835

VERSION

KEYWORDS

SOURCE

ORGANISM

bonnet macaque.

Macaca radiata

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

Cercopithecoidea; Macaca.

REFERENCE

AUTHORS

1 (bases 1 to 322)

Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and

Brahmachari, S.K.

CAG repeat instability at SCA2 locus: anchoring CAA interruptions

and linked single nucleotide polymorphisms

Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

JOURNAL

PUBMED

REFERENCE 2 (bases 1 to 322)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..322
/organism="Macaca radiata"
/db_xref="taxon:9548"
<1..>322
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT 32 a 155 c 95 g 40 t
ORIGIN

Query Match 91.9%; Score 29.4; DB 9; Length 322;
Best Local Similarity 96.8%; Pred. No. 27;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cgcgaaccgcgcctcccgctcgagccgcg 31
|||||
Db 96 CGCCAACCGCGCTCCCTCGCTGCGGCCG 126

RESULT 13
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS Presbytlis entellus SCA2 gene, partial sequence.
DEFINITION AF330030
ACCESSION AF330030
VERSION AF330030.1 GI:12382832
KEYWORDS
SOURCE Hanuman langur,
ORGANISM Presbytlis entellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Colobinae; Presbytis.
1 (bases 1 to 384)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490

REFERENCE
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..384
/organism="Presbytis entellus"
/db_xref="taxon:9574"
<1..>384
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT 46 a 178 c 109 g 51 t
ORIGIN

Query Match 91.9%; Score 29.4; DB 9; Length 384;
Best Local Similarity 96.8%; Pred. No. 26;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 cgcgaaccgcgcctcccgctcgagccgcg 31
|||||
Db 73 CGCCAACCGCGCTCCCTCGCTGCGGCCG 103

RESULT 14
AF330029

LOCUS AF330029 409 bp DNA linear PRI 08-NOV-2001
DEFINITION Gorilla gorilla SCA2 gene, partial sequence.
ACCESSION AF330029
VERSION AF330029.1 GI:12382831
KEYWORDS
SOURCE Gorilla.
ORGANISM Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
1 (bases 1 to 409)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490

REFERENCE
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source Location/Qualifiers
1..409
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
<1..>409
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT 35 a 196 c 120 g 58 t
ORIGIN

Query Match 88.8%; Score 28.4; DB 9; Length 409;
Best Local Similarity 96.7%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gccaaccgcgcctcccgctcgagccgcg 31
|||||
Db 102 GCCAACCGCGCTCCCTCGCTGCGGCCG 131

RESULT 15
SC5E9 29924 bp DNA linear BCT 04-JAN-2001
LOCUS Streptomyces coelicolor cosmid 5E9.
DEFINITION AL446003
ACCESSION AL446003
VERSION AL446003.1 GI:11061544
KEYWORDS
anti-sigma factor antagonist; DNA-binding; hydrolase; IS110;
IS1650; killer toxin-like protein; oxidoreductase; pseudogene;
secreted; transcriptional regulator; transposase.
Streptomyces coelicolor.
Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 29924)
Redenbach,M., Kiser,H.M., Denapalte,D., Elchner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351

JOURNAL
MEDLINE 2 (bases 1 to 29924)
REFERENCE Seeger,K.J. and Harris,D.
AUTHORS unpublished
JOURNAL 3 (bases 1 to 29924)
TITLE Bentley,S.D., Parkhill,J., Barrell,B.G. and Randsdram,M.A.
Direct Submission
JOURNAL Submitted (26-OCT-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5E9 lies between and overlaps cosmids 8D11 and 10B8A on the AseI-A genomic restriction fragment.

FEATURES

```

source
    1. .29924
        /organism="Streptomyces coelicolor"
        /db_xref="taxon:1902"
    1. .29924
        /organism="Streptomyces coelicolor A3(2)"
        /strain="A3(2)"
        /db_xref="taxon:100226"
        /clone="cosmid 5E9"
        complement(1. .279)
        /gene="SC5E9.01c"
    1. .117
        /note="Nominal overlap with Streptomyces coelicolor cosmid 8D11"
        complement(<1. .279)
        /gene="SC5E9.01c"
        /note="SC5E9.01c, unknown, len: 93aa"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CAC14481.1"
        /db_xref="GI:11061545"
        /translation="MPCMAEVPKSHGWTFTNHAHVAIAADNPARIKIDIAHRCRLTERAVORIIISDLDDQGYLSHTRDGRPTNTRYRIEPEKVLHPAEAGLTVAA"
    358. .362
    371. .754
        /gene="SC5E9.02"
    371. .754
        /gene="SC5E9.02"
        /note="SC5E9.02, possible anti-sigma factor antagonist, len: 127aa; weakly similar to many eg. SW:09WYX8 (RSBV_STRCO) anti-sigma B factor antagonist from Streptomyces coelicolor (113 aa) fasta scores: opt: 118, z-score: 164.6, E(): 0.11, 27.7% identity in 94 aa overlap. Contains Pfam match to entry PF01740 STAS, STAS domain."
        /codon_start=1
        /transl_table=11
        /product="putative anti-sigma factor antagonist"
        /protein_id="CAC14482.1"

```

misc-feature

```

/db_xref="GI:11061546"
/translation="MSLHKAVTGTADAVSRPGDQASVLYERGVVYGCGER
DHSITPSQALGTAARHETKVLKASGITFPADBALNLILITGRVSDLRVAAAROL
RRLLEITGVADLAKRSIVSEBAATC"
446. .742
/gene="SC5E9.02"
/note="Pfam match to entry PF01740 STAS, STAS domain, score 38.20, E-value 1.9e-07"
901. .904
912. .1154
/gene="SC5E9.03"
912. .1154
/gene="SC5E9.03"

```

RBS

```

912. .1154
/gene="SC5E9.03"
912. .1154
/gene="SC5E9.03"

```

CDS

```

912. .1154
/gene="SC5E9.03"
/note="SC5E9.03, conserved hypothetical protein, len: 80aa; similar to others from Streptomyces coelicolor eg. TR:054206 (EMBL:AJ001206) pepa hypothetical protein from the glycogen metabolism cluster (90 aa) fasta scores: opt: 116, z-score: 182.1, E(): 0.011, 36.3% identity in 80 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC14483.1"
/db_xref="GI:11061547"
/translation="MTPAEKELRAVLARFQARIDHDVPTGRTSRALDVTYTLGV
ITGARTAEALNTADALLARYDRTSAADDETLAA"
1168. .1172
1175. .1399
/gene="SC5E9.04"
1175. .1399
/gene="SC5E9.04"
/note="SC5E9.04, unknown, len: 74aa"

```

RBS

```

1175. .1399
/gene="SC5E9.04"
1175. .1399
/gene="SC5E9.04"

```

CDS

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC14484.1"
/db_xref="GI:11061548"
/translation="MIAOGATIVGDAQTSVPVLMPLSPRAVTPARSEARRESRAFP
SAVPRHHRVAVIARRAPASSITCFPEVR"
complement(2005. .2883)
/note="Insertion element IS1650"
/label="IS1650"
complement(2026. .2472)
/gene="SC5E9.05c"
complement(2026. .2472)

```

misc-feature

```

/gene="SC5E9.05c"
/note="SC5E9.05c, possible IS1650 transposase, partial CDS, len: 148aa; similar to many, identical to TR:Q9XAE7 (EMBL:AL079356) putative transposase from Streptomyces coelicolor (148 aa). May be translated by frameshift from upstream CDS."
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="CAC14485.1"
/db_xref="GI:11061549"
/translation="MTTKIHLACDGGGRPLAFLITAGVNDCTOFEOVVARIRIORCG
PGRPRPRPERVADKXSSKRTFTYRRGICRAIPERIDOINGRIRRGESLCRIDRA
AYRRRVVERCFKELKHNKALATRYKRRRHVQALVTACLMLP"
complement(2469. .2879)
/gene="SC5E9.06c"
complement(2469. .2879)

```

CDS

```

/gene="SC5E9.06c"
/note="SC5E9.06c, possible IS1650 transposase, partial CDS, len: 136aa; similar to many, identical to TR:Q9XAE6 (EMBL:AL079356) putative transposase from Streptomyces coelicolor (136 aa). Translated may frameshift into downstream CDS. Contains Pfam match to entry PF01511 transposase.6, Transposase."
/codon_start=1
/transl_table=11
/product="putative transposase"

```

gene

CDS

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 22:06:40 ; Search time 906.46 Seconds

(without alignments)
60.611 Million cell updates/sec

Title: US-09-707-919-7

Sequence: 1 cgcacccgcgcctcccgctcgccgcgcct 32

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	516	19	SCA2 gene fragment
2	32	100.0	4200	18	Sphino cerebellar at
3	32	100.0	4367	19	Gene causative of
4	32	100.0	4481	20	Human SCA2 cDNA in
5	32	100.0	4481	20	Human SCA2 DNA. H
6	31.6	98.8	355	19	SCA2 gene fragment
7	31.6	98.8	623	19	SCA2 gene fragment
8	22	68.8	6862	22	Tumour suppressor
9	22	68.8	6862	24	Human immune syste

C	10	22	68.8	6862	24	AA561082	Human gene regulat
C	11	21.4	66.9	98	22	AA67698	Insulator plasmid
C	12	21	65.6	3682	21	AA97998	Human T gene DNA f
C	13	21	65.6	24000	21	AA88551	Human dual-specifi
C	14	20.8	65.0	645	21	AA38874	Human tumour suppr
C	15	20.8	65.0	1308	19	AAV19115	Human secreted apo
C	16	20.8	65.0	2075	22	AAAD17401	Human secreted Frl
C	17	20.8	65.0	2124	20	AAV84394	Partial FRP genom
C	18	20.8	65.0	3717	21	AA64660	DNA encoding centr
C	19	20.8	65.0	4469	22	AA512954	Human Fritzzle Rela
C	20	20.8	65.0	4497	20	AAV84395	Human Fritzzled-rel
C	21	20.8	65.0	4616	22	AAH72901	Human cervical can
C	22	20.6	64.4	727	22	AAH08591	Human cDNA clone (
C	23	20.6	64.4	2858	22	AAH18315	Human cDNA sequenc
C	24	20.4	63.7	232	22	AAH07599	Human cDNA clone (
C	25	20.4	63.7	551	23	AA578287	DNA encoding novel
C	26	20.4	63.7	675	22	AA522161	DNA encoding novel
C	27	20.4	63.7	725	22	AAH05739	Human cDNA clone (
C	28	20.4	63.7	1844	22	AAI93906	Human stomach canc
C	29	20.4	63.7	1844	22	AAH18032	Human cDNA sequenc
C	30	20.4	63.7	2651	22	AAH18567	Human cDNA sequenc
C	31	20.4	62.5	411	21	AA393724	zee mays DNA fragm
C	32	20	62.5	1927	24	AA518807	DNA encoding cance
C	33	20	62.5	2110	24	AA518808	DNA encoding cance
C	34	20	62.5	8034	24	AA518806	DNA encoding cance
C	35	20	62.5	10211	19	AAV62152	HSV-2 strain SB5 C
C	36	20	62.5	117213	19	AAV62176	Human herpesvirus
C	37	20	62.5	154746	24	AA525519	Human secreted pro
C	38	19.8	61.9	146	21	AA510674	Pseudomonas aerugi
C	39	19.8	61.9	279	23	AA548747	Pseudomonas aerugi
C	40	19.8	61.9	407	23	AA547894	Colon tumour relat
C	41	19.8	61.9	407	22	AAI28632	DNA encoding novel
C	42	19.8	61.9	588	23	AA591762	Pseudomonas aerugi
C	43	19.8	61.9	2178	23	AA551488	Human PRO531 nucle
C	44	19.8	61.9	2325	19	AA795400	
C	45	19.8	61.9	2738	20	AA234233	

ALIGNMENTS

RESULT	1
AAV06551	AAV06551 standard; DNA; 516 BP.
ID	XX
AC	XX
XX	AAV06551;
DT	06-JUL-1998 (first entry)
XX	XX
DE	SCA2 gene fragment including CAG repeat region.
XX	XX
XX	SCA2 gene: sphino cerebellar ataxia-2; ataxin-2; human;
KW	diagnosis; olivo-ponto-cerebellar atrophy; ss; ds.
XX	XX
OS	Homo sapiens.
XX	XX
FH	Key
FT	primer_bind
FT	Location/Qualifiers
FT	complement (241..257)
FT	/tag= a
FT	/note= "primer SCA2-A binding site"
FT	/tag= b
FT	/tag= b
FT	/note= "primer SCA2-B binding site"
FT	/tag= C
FT	/note= "predicted splice site"
FT	/tag= d
FT	/note= "CAG repeat region"
FT	/tag= e
FT	/note= "CAG repeat"
FT	repeat_unit
FT	repeat_unit
FT	270..272


```

FT      /*tag= d
FT      /note= "putative Kozak consensus signal"
FT      258..323
FT      /tag= e
FT      /note= "encodes polyglutamine repeat region: contains
FT      repeat_unit
FT      258..260
FT      /tag= f
FT      /note= "CAG repeats"
FT      1..3986
FT      /tag= g
FT      /note= "sequence contained in DAN1 clone"
FT      3987..4200
FT      /tag= h
FT      /note= "derived from the EST's AAN92640, AAN90240 and
FT      AAI3574 from dbEST database"
FT      misc_feature
FT      4023..4029
FT      /tag= i
FT      /note= "region which differs in length between the
FT      sequences of the EST clones AAN92640, AAN90240
FT      and AAI3574"
FT
FT      WO9717445-A1.
FT
FT      15-MAY-1997.
FT
FT      08-NOV-1996; 96WO-FR01773.
FT
FT      10-NOV-1995; 95FR-0013576.
FT
FT      (CNRS ) CNRS CENT NAT RECH SCI.
FT      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
FT      Lutz Y, Mandel J, Tora L, Trollier Y;
FT
FT      WPI: 1997-281034/25.
FT      P-PSDB; AAM24800, AAM24801.
FT
FT      Antibody 1c2 used for treating or preventing neuro-degenerative
FT      diseases - associated with proteins containing long poly:glutamine
FT      repeats, e.g. Huntington's disease
FT
FT      Claim 21; Page 45-47; 69pp; French.
FT
FT      The invention relates to a monoclonal antibody (MAB) 1C2 for the
FT      treatment of neurodegenerative diseases associated with the presence
FT      of polyglutamine repeat regions. This MAB is already known for its
FT      affinity to the YATA binding protein (TBP) transcription initiation
FT      factor, especially at the amino acid sequence LEQQRQ000Q found at
FT      the N-terminus of TBP. MAB 1C2 has been shown to have a high affinity
FT      for polyglutamine repeats with a proportional affinity to the number
FT      of glutamine repeats. This affinity has been used to identify genes
FT      encoding proteins containing long polyglutamine repeats which are
FT      implicated in neurodegenerative diseases. A screen of an expression
FT      library, generated from a lymphoblastic cell line from a patient
FT      suffering from spinocerebellar ataxia (SCA), with MAB 1C2 isolated 6
FT      new sequences (AA178906-178911) encoding polyglutamine repeats. MAB 1C2
FT      also isolated the complete SCA2 gene in clone DAN1 (sequence presented
FT      here). The sequence appears to contain 2 open reading frames (ORF) the
FT      second of which may be generated by an frameshift slippage or by an
FT      alternative splicing event. The first ORF also encodes a 22 amino acid
FT      polyglutamine repeat region near the N-terminus of the protein. Normal
FT      SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
FT      interspersed whereas the mutant sequence from patients with SCA
FT      contains at least 30, preferably 37-50 CAG repeats.
FT
FT      MAB 1C2, active fragment of it or nucleic acids encoding it are
FT      specifically used to treat Huntington's disease, SCA types 1-5 or 7,
FT      X-linked spino-bulbar muscular atrophy (Kennedy disease),
FT      dentrocrubral-pallidolusial atrophy, dominant autosomal spinocerebellar
FT      ataxia, familial spastic paraplegia, bipolar affective disorder, manic
FT      depressive psychoses and schizophrenia.
FT
FT      Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

```

```

OY      1 cgcacaccgcgcctcccgctcggcgcgcgt 32
DB      121 cgcacaccgcgcctcccgctcggcgcgcgt 152

RESULT 3
AAV30270
ID      AAV30270 standard; DNA; 4367 BP.
AC      AAV30270;
DT      02-OCT-1998 (first entry)
DE      Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
XX
XX      Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
XX      CAG repeat; neurodegenerative disease; ds.
XX
XX      Homo sapiens.
XX
XX      Key      Location/Qualifiers
XX      CDS      49..3990
XX      FT      /*tag= a
XX      FT      /product= "Spinocerebellar ataxia type 2 associated
XX      FT      protein"
XX      repeat_region 544..612
XX      FT      /*tag= b
XX      FT      /note= "normal CAG repeat region; this is increased in
XX      FT      patients with SCA2"
XX      repeat_unit 544..546
XX      FT      /*tag= c
XX
XX      WO9818920-A1.
XX
XX      07-MAY-1998.
XX
XX      30-OCT-1997; 97WO-JP03946.
XX
XX      30-OCT-1996; 96JP-0304059.
XX
XX      (SRLS-) SRL INC.
XX
XX      Sempel K, Tsuji S;
XX
XX      WPI: 1998-272215/24.
XX      P-PSDB; AAM60213.
XX
XX      Nucleic acid fragments associated with spinocerebellar ataxia type 2
XX      - contain increased number of CAG repeat region compared to normal
XX      gene
XX
XX      Claim 1; Pages 13-22; 38pp; Japanese.
XX
XX      This represents the sequence of a gene causative of spinocerebellar
XX      ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
XX      with SCA2, has a tri-nucleotide (CAG) repeat region which in the
XX      expression product produces a polyglutamine sequence from Gln-166 to
XX      Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
XX      patients this number is increased to 35-100. Peptides encoded by nucleic
XX      acid fragments (DNA or RNA) containing sequences from the SCA2 associated
XX      gene, antibodies recognising the peptides and antisense nucleic acids
XX      hybridising with the nucleic acid fragments can be used for the
XX      investigation and diagnosis of SCA2. They can also be used for the
XX      treatment of SCA2 by antisense therapy or gene therapy.
XX
XX      Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

```


SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 32; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacaaccgcgcctcccgctcgcgcgcgt 32
|||||
DB 521 cgcacaaccgcgcctcccgctcgcgcgcgt 552

RESULT 5

AA223428
ID AA223428 standard; DNA; 4481 BP.

AC AA223428;

DT 19-JAN-2000 (first entry)

DE Human SCA2 DNA.

XX Proapoptotic; dependence domain; P75NTR; androgen receptor; DCC;
KM huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KM atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KM Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KM dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KM neoplastic; malignant; autoimmune; fibrotic; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 163..4101

FT /*tag= a

FT /product= "SCA2"

XX WO9945944-A1.

XX 16-SEP-1999.

XX 11-MAR-1999; 99WO-US05250.

XX 12-MAR-1998; 98US-0041886.

XX (BURN-) BURNHAM INST.

XX Bredesen DE, Rabizadeh S;

XX MPI: 1999-561617/47.

XX P-PSDB: AAY33495.

XX New proapoptotic dependence peptides, used to develop products for
XX treating, e.g. Alzheimer's disease -

XX Disclosure: Page 130-135; 199pp; English.

CC This invention describes novel pure proapoptotic dependence peptides
CC which comprise a sequence of an active dependence domain selected from
CC dependence polypeptides consisting of P75NTR, androgen receptor, DCC,
CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC of inducing cell death and can be used to develop products to mediate or
CC inhibit apoptosis. The methods can be used for reducing the severity of
CC a proapoptotic dependence domain mediated pathological conditions e.g.
CC Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC spinocerebellar ataxias, dentatorubropallidoluysian atrophy,
CC Machado-Joseph disease, stroke or head trauma. They can also be used for
CC reducing the severity of a pathological condition mediated by upregulated
CC cell proliferation or cell survival e.g. neoplastic, malignant,
CC autoimmune or fibrotic conditions. This sequence encodes the human
CC SCA2 polypeptide described in the method of the invention.

SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 32; DB 20; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacaaccgcgcctcccgctcgcgcgcgt 32
|||||
DB 521 cgcacaaccgcgcctcccgctcgcgcgcgt 552

RESULT 6

AAV17224
ID AAV17224 standard; DNA; 355 BP.

AC AAV17224;

DT 29-JUN-1998 (first entry)

DE SCA2 gene fragment.

XX SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 341..355

FT /*tag= a

FT /note= "SCA2 protein fragment"

XX WO9803679-A1.

XX 29-JAN-1998.

XX 18-JUL-1996; 96WO-JP01999.

XX 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sempel K, Tsuji S;

XX MPI: 1998-120796/11.

XX P-PSDB: AAM41370.

XX Diagnosing spinocerebellar ataxis type II - by PCR and determining
XX number of CAG repeat units

XX Claim 1; Page 10; 23pp; Japanese.

CC This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxis type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxis type II.

XX Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

Query Match 98.8%; Score 31.6; DB 19; Length 355;
Best Local Similarity 96.9%; Pred. No. 0.11;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacaaccgcgcctcccgctcgcgcgcgt 32
|||||
DB 219 cgcacaaccgcgcctcccgctcgcgcgcgt 250

RESULT 7

AAV17229
ID AAV17229 standard; DNA; 623 BP.

XX

```

AC  AAV17229;
XX
XX  29-JUN-1998 (first entry)
XX
XX  SCA2 gene fragment.
XX
XX  SCA2 gene: spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX
XX  Synthetic.
XX
XX  Key      Location/Qualifiers
XX  CDS      341..583
XX           /*tag=
XX           /note= "SCA2 protein fragment, no stop codon given"
XX
XX  WO9803679-A1.
XX
XX  29-JAN-1998.
XX
XX  18-JUL-1996; 96WO-JP01999.
XX
XX  18-JUL-1996; 96WO-JP01999.
XX
XX  18-JUL-1996; 96WO-JP01999.
XX
XX  (SRLS-) SRL INC.
XX
XX  Saepel K, Tsuji S;
XX
XX  WPI: 1998-120796/11.
XX
XX  P-PSDB; AAM41372.
XX
XX  Diagnosing spinocerebellar ataxis type II - by PCR and determining
XX  number of CAG repeat units
XX
XX  Example 1; Page 11-12; 23pp; Japanese.
XX
XX  This sequence represents a fragment of the SCA2 gene. It can be used in
XX  the method of the invention for diagnosing spinocerebellar ataxis type
XX  II, by performing PCR on the test DNA using two primers hybridising to
XX  parts of the SCA2 gene sequence, and determining the number of CAG
XX  repeats in the amplified products. The method provides an easy means for
XX  the diagnosis of spinocerebellar ataxis type II.
XX
XX  Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;
XX
XX  Query Match      98.8%; Score 31.6; DB 19; Length 623;
XX  Best Local Similarity 96.9%; Pred. No. 0.1;
XX  Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 cgcacaaccgcgcctcccgctcgagccgcgt 32
XX  |||||||||||||||||||||||||||||
XX  219 cgcacaaccgcgcctcccgctcgagccgcgy 250
XX
XX  RESULT 8
XX  ID AAS46300/C
XX  AAS46300 standard; DNA; 6862 BP.
XX
XX  AAS46300;
XX
XX  18-DEC-2001 (first entry)
XX
XX  Tumour suppressor gene derived chemically modified sequence #22.
XX
XX  Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX  cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
XX  cytosine methylation; ds.
XX
XX  Homo sapiens.
XX
XX  WO200168912-A2.
XX
XX  20-SEP-2001.
XX

```

```

XX
XX  15-MAR-2001; 2001WO-EP02955.
XX
XX  15-MAR-2000; 2000DE-1013847.
XX
XX  06-APR-2000; 2000DE-1019058.
XX
XX  07-APR-2000; 2000DE-1019173.
XX
XX  30-JUN-2000; 2000DE-1032529.
XX
XX  01-SEP-2000; 2000DE-1043826.
XX
XX  (EPIC-) EPIGENOMICS AG.
XX
XX  Olek A, Piepenbrock C, Berlin K;
XX
XX  WPI: 2001-602752/68.
XX
XX  Fragments of chemically modified genes associated with tumour suppressor
XX  genes and oncogenes, useful in designing primers and probes for
XX  analysing diseases associated with cytosine methylation state e.g.
XX  cancer
XX
XX  Claim 1; SEQ ID NO 22; 27pp; English.
XX
XX  The invention relates to a nucleic acid comprising a sequence of 18
XX  bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX  bisulphite, of genes associated with tumour suppression and
XX  oncogenes having a sequence taken from 536 (actually 533 since
XX  numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX  (SS) and sequences complementary to (SS). The nucleic acid may be a
XX  peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX  form part of a set of probes for detecting the cytosine methylation state
XX  and/or single nucleotide polymorphisms and also to be used in an
XX  array for analysing diseases associated with Cpg dinucleotides e.g.
XX  cancers and tumours. The probes can also be used in a method for
XX  ascertaining genetic and/or epigenetic parameters for the diagnosis
XX  and/or therapy of existing diseases or the predisposition to specific
XX  diseases, by analysing cytosine methylations. The parameters may be
XX  compared to another set of genetic and/or epigenetic parameters, the
XX  differences serving as basis for diagnosis and/or prognosis events which
XX  are disadvantageous to patients. The present sequence is one of the
XX  533 genomic sequences derived from tumour suppressor genes and
XX  oncogenes. Sequences with even numbered Seq ID numbers are the
XX  complementary sequence of the corresponding odd numbered sequence (e.g.
XX  CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
XX  is missing).
XX  Note: The sequence data for this patent did not form part
XX  of the printed specification, but was obtained in electronic
XX  format directly from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other;
XX
XX  Query Match      68.8%; Score 22; DB 22; Length 6862;
XX  Best Local Similarity 83.3%; Pred. No. 79;
XX  Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX  1 cgcacaaccgcgcctcccgctcgagccgc 30
XX  ||||||||||||| || ||| || |||||
XX  278 CGACAACCGCGCGCGCGCGCGCACGCCGCC 249
XX
XX  RESULT 9
XX  ID ABL32223/C
XX  ABL32223 standard; DNA; 6862 BP.
XX
XX  ABL32223;
XX
XX  26-MAR-2002 (first entry)
XX
XX  Human immune system associated gene SEQ ID NO: 196.
XX
XX  Human; immune system disease; cytosine methylation; antiasthmatic;
XX  antiarteriosclerotic; antiataemic; cytostatic; nootropic;
XX
XX

```

KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; anarthritic; antidiabetic; antipsorific;
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KM gene; ds.
 OS Homo sapiens.
 XX WO200200928-A2.
 PN
 XX
 PD 03-JAN-2002.
 PF
 XX 02-JUL-2001: 2001WO-EP07537.
 XX
 PR 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI: 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1: SEQ ID NO 196; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SO Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other;
 OY
 1 cggcaaccgagcctcccgctcgagcc 30
 || ||||| ||||| || |||||
 Db 278 CGACACCGCCGCCGCCGCCGCCGCC 249
 RESULT 10
 AAS61082/c
 ID AAS61082 standard; DNA: 6862 BP.
 XX
 AC AAS61082;
 XX
 DT 29-JAN-2002 (first entry)
 DE Human gene regulation-associated gene oligonucleotide #37.
 XX
 KM Human: Gene regulation-associated gene; severe combined immunodeficiency;
 KM cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KM asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KM renal disease; Preeclampsia; cardiac allograft vascular disease;
 KM colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KM immunostimulant; cardiact; antiinflammatory; coagulant; antiasthmatic;
 KM nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX

PD 18-OCT-2001.
 XX
 XX 06-APR-2001: 2001WO-EP03968.
 PF
 XX 06-APR-2000: 2000DE-1019058.
 PR 07-APR-2000: 2000DE-1019173.
 PR 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI: 2002-017470/02.
 DR
 XX New nucleic acid sequences from chemically modified genes associated
 PT with gene regulation, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
 PT disease -
 PS Claim 1: SEQ ID NO 38; 26pp; English.
 XX
 CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The
 CC chemical pretreatment converts cytosine bases unmethylated at the
 CC 5-position to uracil or another base with hybridisation behaviour
 CC dissimilar to cytosine, to enable analysis of cytosine methylations.
 CC The DNA sequences, oligomers (or sets/arrays) and method are
 CC useful in the diagnosis of diseases (or predisposition to diseases)
 CC associated with gene regulation and in therapy of such diseases, by
 CC enabling analysis of the cytosine methylation patterns of such genes,
 CC kits are provided. They are especially useful in diagnosis
 CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
 CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
 CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
 CC preeclampsia, graft versus-host disease. The present sequence is a
 CC sequence included in the sequence data for this specification and is
 CC associated with the human gene regulation-associated genes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC http://wipo.int/pub/published_pct_sequences
 XX
 SO Sequence 6862 BP; 1370 A; 518 C; 2038 G; 2936 T; 0 other;
 OY
 1 cggcaaccgagcctcccgctcgagcc 30
 || ||||| ||||| || |||||
 Db 278 CGACACCGCCGCCGCCGCCGCCGCCGCC 249
 RESULT 11
 AAF67698/c
 ID AAF67698 standard; DNA: 98 BP.
 XX
 AC AAF67698;
 XX
 DT 12-APR-2001 (first entry)
 DE Insulator plasmid enhancer blocking sequence Apb SEQ ID NO: 56.
 XX
 KM Chicken; human; insulator; enhancer; DNA binding protein;
 KM gene expression; gene therapy; insulin-like growth factor-2; Igf2;
 KM knockout mouse; ds.
 XX
 OS Unidentified.
 XX
 PN WO200102553-A2.
 XX

XX 11-JAN-2001.
XX 19-APR-2000; 2000MO-US10509.
XX 30-JUN-1999; 990US-0141728.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Bell AC, West AC, Felsenfeld G;
XX WPI; 2001-091803/10.
XX Isolated DNA molecule useful for the regulation of gene expression and
XX function in mammals and plants -
XX Example 1; Page 65; 96pp; English.
XX The present invention provides the sequence of a enhancer-blocking
XX insulator from the chicken. Also provided are insulators 1 om the murine,
XX rat and human insulin-like growth factor-2 (Igf2) genes. The insulators
XX can be used to modulate gene expression, for example in gene therapy and
XX in knockout mouse production.
XX Sequence 98 BP; 7 A; 37 C; 43 G; 11 T; 0 other;
SQ

Query Match 66.9%; Score 21.4; DB 22; Length 98;
Best Local Similarity 80.6%; Pred. No. 2.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgcgcgcg 31
||||| ||||| ||||| ||||| |||||
DB 94 CGCCTATCCGGCCACCCGCTCGCACCCG 64

RESULT 12
AAA97998/C
ID AAA97998 standard; DNA; 3682 BP.
XX
AC AAA97998;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human T gene DNA fragment #1.
XX
KW T gene; human; central nervous system development; CNS; neurotropic;
KW neuroleptic; antidepressant; gene therapy; antisense; treatment;
KW schizophrenia; autism; manic depression; mental retardation; ds.
XX
OS Homo sapiens.
XX
PN DE19908423-A1.
XX
PD 31-AUG-2000.
XX
PE 26-FEB-1999; 99DE-1008423.
XX
PR 26-FEB-1999; 99DE-1008423.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
PI Poustka A, Coy J;
XX
PI WPI; 2000-580150/55.
XX
DR DNA encoding a protein involved in development of the central nervous
PT system (CNS); antisense sequences, ribozymes and antibodies, useful for
PT treatment of, e.g. schizophrenia and manic depression -
XX
PS Claim 1a; Fig 3; 86pp; German.
XX
CC This invention describes a novel DNA sequence, which encodes a protein

CC that is involved in development of the central nervous system (CNS) and
CC has tissue and development-specific expression. The products of the
CC invention have neurotropic, neuroleptic and antidepressant activity and can
CC be used for gene therapy and antisense inhibition. The method also
CC describes a method for producing (1) antisense RNA that is complementary
CC to DNA as above, which can reduce or inhibit synthesis of the protein
CC coding DNA; (2) a ribozyme, which is complementary to DNA as above, which
CC specifically binds to and cleaves transcribed DNA, which reduces or
CC inhibits synthesis of the protein coding DNA; (3) an expression vector,
CC containing DNA as above, or which encodes antisense RNA or a ribozyme;
CC (4) a host cell transformed with a vector as in (3); (5) a protein,
CC encoded by DNA as above; (6) a method to produce the protein of (5)
CC comprising culturing the cell of (4) and isolating the protein from the
CC cell or the culture medium; (7) an antibody targeted against the protein
CC of (5); (8) a diagnostic method to detect disturbed expression of the
CC protein of (5) or to detect altered forms of the protein by contacting a
CC sample with a DNA sequence or antibody and determining direct or indirect
CC contact, and comparing the expression of the protein with a healthy
CC patient; (9) a diagnostic kit to perform the method of (8); (10) a
CC non-human transgenic animal, where the naturally occurring T gene has an
CC altered gene structure or sequence; and (11) a method to produce a
CC non-human animal as in (10). The DNA, derived from the T gene encodes a T
CC protein (TP) which is involved in development of the central nervous
CC system. Antisense sequences, ribozymes and antibodies are useful for
CC treatment of disorders of the CNS including schizophrenia, autism, manic
CC depression and mental retardation. This sequence encodes a fragment of
CC the human T protein described in the method of the invention.
XX
SQ Sequence 3682 BP; 869 A; 888 C; 933 G; 992 T; 0 other;
XX

Query Match 65.6%; Score 21; DB 21; Length 3682;
Best Local Similarity 82.8%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgcgcgcg 29
||||| ||||| ||||| ||||| |||||
DB 2263 CGCCGGCGCGCGCTCCCGCCCGCGCC 2235

RESULT 13
AAA88551
ID AAA88551 standard; DNA; 24000 BP.
XX
AC AAA88551;
XX
DT 22-JAN-2001 (first entry)
XX
DE Human dual-specificity phosphatase-1 (DSP-1) gene.
XX
KW DSP-1; dual-specificity phosphatase-1; human; cell proliferation;
KW cell differentiation; cell survival; cell cycle; dephosphorylation;
KW signal transduction; MAP-kinase; cancer; graft versus host disease;
KW allergy; autoimmune disease; metabolic disease; therapy;
XX
XX chromosome 17; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
XX
FT exon 42..109
FT /*tag- a
FT /number- 1
FT 110..20823
FT /*tag- b
FT /number- 1
FT 20824..20911
FT /*tag- c
FT /number- 2
FT 20824..21034
FT /*tag- d
FT /number- 2a
FT /note- "Alternative, extended version of exon 2"
FT 20912..22327
FT Intron

```

FT      /tag= e
FT      /number= 2
FT      21035..22327
FT      Intron
FT      /tag= f
FT      /number= 2a
FT      22328..23309
FT      exon
FT      /tag= g
FT      /number= 3
FT      22420..23016
FT      CDS
FT      /tag= h
PN      WO200053636-A2.
PD      14-SEP-2000.
PF      08-MAR-2000; 2000MO-US06154.
PR      08-MAR-1999; 99US-0123255.
PA      (CEPT-) CEPTR INC.
PI      Lucbe RM, Wei B;
PT      WPI: 2000-579365/54.
PT      P-PSDB; AAB19602.
PS      New isolated polypeptide having the sequence of dual-specificity
PS      phosphatase-1 (DSP-1) is useful for treating a patient with a disorder
PS      associated with DSP-1 activity e.g. cancer and autoimmune diseases -
XX      Example 1; Fig 3A-J; 74pp; English.
XX      The present sequence is that of the human dual-specificity
XX      phosphatase-1 (DSP-1) gene on chromosome 17. The gene was
XX      identified in genomic sequences obtained from an expressed sequence
XX      tag database screened with a conserved motif (see AAB19604) of known
XX      DSPs. DSP-1 dephosphorylates both phosphothreonine/serine and
XX      phosphotyrosine residues in DSP-1 substrates such as activated
XX      mitogen-activated protein kinase (MAP-kinase). DSP-1 has sequence
XX      homology to other MAP-kinase phosphatases. It is expressed at high
XX      levels in the human heart, testis and liver, and at lower levels in
XX      other tissues. Methods are provided for recombinant production of
XX      DSP-1 polypeptides, and for using DSP-1 polypeptides, antibodies
XX      and polynucleotides to detect DSP-1 expression, to screen for agents
XX      that modulate DSP-1 activity e.g. within a combinatorial library,
XX      and for using such agents to modulate cell proliferation, cell
XX      differentiation or cell survival, through modulation of pattern of
XX      gene expression, apoptosis or cell cycle. In particular, the cell
XX      displays contact inhibition of cell growth, anchorage-dependent
XX      growth or an altered intercellular adhesion property, or is a cell
XX      present in a patient afflicted with a disorder associated with
XX      DSP-1 activity, such as cancer, graft-versus host disease,
XX      autoimmune disease, allergy, metabolic disease, abnormal cell
XX      growth, abnormal cell proliferation and abnormal cell cycle.
XX      Sequence 24000 BP; 6038 A; 5505 C; 5602 G; 6855 T; 0 other;
SQ

```

Query Match 65.6%; Score 21; DB 21; Length 24000;
 Best Local Similarity 82.8%; Pred. No. 1.4e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY      4 caaccgcgcctcccgctcgagccgct 32
OY      ||||| ||||| ||||| ||||| ||
DB      83 caaccgcgcgcgcgcgcgcgcgcgct 111

```

RESULT 14
 AAC58874/c
 ID AAC58874 standard; DNA; 645 BP.
 XX
 AC AAC58874;
 XX

```

DT      25-JAN-2001 (first entry)
XX      Human tumour suppressor BRG1 gene exon 1.
DE      Human: BRG1; tumour suppressor gene; cancer; chromosome 19p13.1;
XX      retinoblastoma tumour suppressor gene; RB; drug screening; gene therapy;
XX      drug design; peptide therapy; animal model; ss.
XX      Homo sapiens.
XX      WO200056931-A1.
XX      28-SEP-2000.
PD      23-MAR-2000; 2000MO-US07678.
PF      23-MAR-1999; 99US-0125806.
PR      (MYRI-) MYRIAD GENETICS INC.
PA      Wong AKC, Tavtigian SV, Teng DH;
PI      WPI: 2000-587668/55.
PT      Diagnosing a polymorphism associated with predisposition for cancer in
PT      humans by determining whether there is a germline alteration of a BRG1
PT      gene or its expression products -
PS      Claim 18; Page 95; 215pp; English.
XX      The present invention is concerned with the use of the human tumour
XX      suppressor gene BRG1 in cancer diagnosis and therapy. This gene is
XX      comprised of several exons, shown in AAC58874-C58903, and has several
XX      splice variants, given in AAC58906-C58912. The protein sequences for
XX      these are shown in AA27552-B27558. BRG1 is a homologue of the Drosophila
XX      protein Dharma, and has been shown to be bound to retinoblastoma tumour
XX      suppressor protein Rb. The BRG1 coding sequence and protein can be used
XX      in the diagnosis and treatment of cancer (for example by gene therapy),
XX      particularly prostate cancer, to identify drugs useful in the treatment
XX      of cancer and in the production of animal models for cancer.
XX      Sequence 645 BP; 68 A; 201 C; 320 G; 56 T; 0 other;
SQ

```

Query Match 65.0%; Score 20.8; DB 21; Length 645;
 Best Local Similarity 78.1%; Pred. No. 2.5e+02;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

OY      1 cgcacccgcgcctcccgctcgagccgct 32
OY      ||||| ||||| ||||| ||||| ||
DB      61 CGCCAGCCGCCCTTCGCTCCGCCGCCGCGCT 30

```

RESULT 15
 AAV19115
 ID AAV19115 standard; DNA; 1308 BP.
 XX
 AC AAV19115;
 XX
 DT 28-AUG-1998 (first entry)
 XX
 DE Human secreted apoptosis-related protein hSARP2 DNA.
 XX
 KW Secreted apoptosis-related protein; SARP; hSARP2; human;
 KW prostate cancer; breast cancer; diagnosis; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 302..1246
 FT /*tag= a
 XX
 PN WO9813493-A2.

OY 1 cgcacccgcgcctcccgctcgagcccgct 32
|||||
DB 521 CGCCACCGCGGCTCCCGCTCGCGCCGCT 552

RESULT 2
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043, 303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
US-09-043-303-1

Query Match 98.8%; Score 31.6; DB 4; Length 355;
Best Local Similarity 96.9%; Pred. No. 0.015;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgagcccgct 32
|||||
DB 219 cgcacccgcgcctcccgctcgagcccgct 250

RESULT 3
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazuhiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043, 303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(583)
; FEATURE:
; OTHER INFORMATION: TSP-2
US-09-043-303-5

Query Match 98.8%; Score 31.6; DB 4; Length 623;
Best Local Similarity 96.9%; Pred. No. 0.015;
Matches 31; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgagcccgct 32

DB 219 cgcacccgcgcctcccgctcgagcccgct 250
|||||

RESULT 4
US-09-387-212-9
; Sequence 9, Application US/09387212A
; Patent No. 6309849
; GENERAL INFORMATION:
; APPLICANT: ROBISON, KEITH E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
; FILE REFERENCE: MNT-090
; CURRENT APPLICATION NUMBER: US/09/387, 212A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-387-212-9

Query Match 61.9%; Score 19.8; DB 4; Length 3001;
Best Local Similarity 77.4%; Pred. No. .73;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgagcccgct 31
|||||
DB 33 cgcctcccgcccgcccgcccgagcccgct 63

RESULT 5
US-08-308-881-5
; Sequence 5, Application US/08308881
; Patent No. 5783672
; GENERAL INFORMATION:
; APPLICANT: Mosley, Bruce
; APPLICANT: Cosman, David J.
; TITLE OF INVENTION: Receptor for Oncostatin M
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple Macintosh
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308, 881
; FILING DATE: 12-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/249, 553
; FILING DATE: 26-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2614-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4171 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: huOSM-Ra
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 368..448
FEATURE:
NAME/KEY: CDS
LOCATION: 368..3307
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 449..3304
US-08-308-881-5

Query Match 61.9%; Score 19.8; DB 1; Length 411;
Best Local Similarity 77.4%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cgcacaccgcgcctcccgctcgagcccg 31
DB 102 CCCGACCGCCGCTCCGCTGCTCGCG 132

RESULT 6
US-09-058-263-5
Sequence 5, Application US/09058263
Patent No. 5891997
GENERAL INFORMATION:
APPLICANT: Mosley, Bruce
APPLICANT: Cosman, David J.
TITLE OF INVENTION: Receptor for Oncostatin M
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,263
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/308,881
FILING DATE: 12-SEP-1994
APPLICATION NUMBER: US 08/249,553
FILING DATE: 26-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2614-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA

HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: huOSM-Ra
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 368..448
FEATURE:
NAME/KEY: CDS
LOCATION: 368..3307
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 449..3304
US-09-058-263-5

Query Match 61.9%; Score 19.8; DB 2; Length 4171;
Best Local Similarity 77.4%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cgcacaccgcgcctcccgctcgagcccg 31
DB 102 CCCGACCGCCGCTCCGCTGCTCGCG 132

RESULT 7
US-09-059-099-5
Sequence 5, Application US/09059099
Patent No. 5925740
GENERAL INFORMATION:
APPLICANT: Mosley, Bruce
APPLICANT: Cosman, David J.
TITLE OF INVENTION: Receptor for Oncostatin M
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,099
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/308,881
FILING DATE: 12-SEP-1994
APPLICATION NUMBER: US 08/249,553
FILING DATE: 26-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2614-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 449..3304
PCT-US95-06530-5

Query Match 61.9%; Score 19.8; DB 5; Length 4171;
Best Local Similarity 77.4%; Pred. No. 71;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 cgcacccgcgcctcccgctcgagccgc 31
DB 102 CCCGACCGCCGCTCCGCTGCTCGCG 132

RESULT 10
US-09-045-973-6
Sequence 6, Application US/09045973
Patent No. 615767
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah
TITLE OF INVENTION: PROTEIN PHOSPHATASE RELATED MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,973
FILING DATE: Filed Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0491 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1729 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT16
CLONE: 3041794
US-09-045-973-6

Query Match 61.3%; Score 19.6; DB 4; Length 1729;
Best Local Similarity 84.6%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 4 caaccgcgcctcccgctcgagcc 29
DB 285 CAACCCGCGCGCGCGCGCGCGCC 310

RESULT 11
US-08-644-271-31
Sequence 31, Application US/08644271
Patent No. 5814478

GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:

INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA

FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..1476
OTHER INFORMATION:
NAME/KEY: Human Agrin
LOCATION: 1..1479
OTHER INFORMATION:
US-08-644-271-31

Query Match 60.0%; Score 19.2; DB 1; Length 1479;
Best Local Similarity 87.5%; Pred. No. 1,2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 accgcgcctcccgctcgagcc 29
DB 1264 ACCGCTCTCTCCGCTGCGCGCC 1287

RESULT 12
US-09-428-517-1/c
Sequence 1, Application US/09428517
Patent No. 6251636
GENERAL INFORMATION:
APPLICANT: Betlach, Mary C.
APPLICANT: Shah, Sanjay Krishnakant
APPLICANT: McDaniel, Robert
APPLICANT: Tang, Li

```

1 TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
2
3 FILE REFERENCE: 30062-20029..00
4
5 CURRENT APPLICATION NUMBER: US/09/428,517
6
7 CURRENT FILING DATE: 1999-10-28
8
9 EARLIER APPLICATION NUMBER: 60/120,254
10
11 EARLIER FILING DATE: 1999-02-16
12
13 EARLIER APPLICATION NUMBER: 60/106,100
14
15 EARLIER FILING DATE: 1998-10-29
16
17 NUMBER OF SEQ ID NOS: 12
18
19 SOFTWARE: Patentln Ver. 2.1
20
21 SEQ ID NO 1
22
23 LENGTH: 50937
24
25 TYPE: DNA
26
27 ORGANISM: Artificial Sequence
28
29 FEATURE:
30
31 OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
32
33 US-09-428-517-1

```

Query Match	59.4%	Score 19	DB 4	Length 50937
Best Local Similarity	81.5%	Pred. No.	1e+02	
Matches 22	Conservative 0	Mismatches 5	Indels 0	Gaps 0

QY	1	cgcaaccgcgcgtcccgctgcgcg	27
Db	39999	CGTCACACGACCTCACCGCTGGCG	39973

RESULT 13
US-08-483-488-5/c
Sequence 5, Application US/08483488
Patent No. 5853985
GENERAL INFORMATION:
APPLICANT: Salbaum, Johannes; Masters, Colin;
APPLICANT: Beyreuther, Konrad
TITLE OF INVENTION: Promoter of the Gene for the
TITLE OF INVENTION: Human Precursor of the Alzheimer's
TITLE OF INVENTION: Disease and its Use
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB
MEDIUM TYPE: Storage
COMPUTER: NEC Powermate SX/20
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,488
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/325,745
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,546
FILING DATE: 16-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,330
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/393,360
FILING DATE: 14-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/385,758
FILING DATE: 26-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 8820450.8

```

? FILING DATE: 30-AUG-1988
? ATTORNEY/AGENT INFORMATION:
? NAME: Kurt G. Briscoe
? REGISTRATION NUMBER: 33,141
? REFERENCE/DOCKET NUMBER: MTI 212.6-KGB
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (914) 332-1700
? TELEFAX: (914) 332-1844
?
? TELEX:
?
? INFORMATION FOR SEQ ID NO: 5:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3804 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
?
US-08-483-488-5

```

Query Match	58.8%	Score 18.8;	DB 2;	Length 3804;
Best Local Similarity	76.7%	Pred. No. 1.5e+02;		
Matches 23;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;

```

QY      2  gccaacccgcgcctcccgctcgcgcccg  31
          || | ||| || ||||| |||| ||||
Db      3654 GCTGATCCGGGCCACCCGCTGGGCACCCG  3625

```

RESULT 4
US-09-082-092-9/c
Sequence 9, Application US/09082092
Patent No. 6251628
GENERAL INFORMATION:
APPLICANT: Nakao, Asumiko
APPLICANT: Moren, Anita
APPLICANT: Heuchel, Rainer
APPLICANT: Itoh, Susumu
APPLICANT: Afrakhte, Mozghan
APPLICANT: Soucheinytskyi, Serhiy
APPLICANT: Brodin, Greger
APPLICANT: Landstrom, Marene
APPLICANT: Heldin, Nils-Erik
APPLICANT: Heldin, Carl-Henrik
APPLICANT: ten Dijke, Peter
TITLE OF INVENTION: SMAD7 AND USES THEREOF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTESTO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082.092
FILING DATE: 20-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/047,221
FILING DATE: 20-MAY-1997
APPLICATION NUMBER: 60/060,465
FILING DATE: 30-SEP-1997
APPLICATION NUMBER: 60/075,940
FILING DATE: 25-FEB-1998
APPLICATION NUMBER: 60/077,033
FILING DATE: 06-MAR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212

```

; REFERENCE/DOCKET NUMBER: L0461/7032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-082-092-9

Query Match          58.1%; Score 18.6; DB 4; Length 1491;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gccaacccgcgcctcccgctcggc 26
   |||| |||| |||| |||| |||| ||
DB 346 gccagcccgccgctcccgctccgc 322

RESULT 15
US-09-288-292A-45/c
; Sequence 45, Application US/09288292A
; Patent No. 6359194
; GENERAL INFORMATION:
; APPLICANT: Dean A. Falb
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
; APPLICANT: Michael A. Gimbirone, Jr.
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of
; FILE REFERENCE: 7853-140-999
; CURRENT FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: US/09/288,292A
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1817
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-288-292A-45

Query Match          58.1%; Score 18.6; DB 4; Length 1817;
Best Local Similarity 84.0%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gccaacccgcgcctcccgctcggc 26
   |||| |||| |||| |||| |||| ||
DB 500 gccagcccgccgctcccgctccgc 476
```

Search completed: August 14, 2002, 21:55:34
Job time: 13767 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:29 : Search time 7749.14 Seconds
(without alignments)
57.736 Million cell updates/sec

Title: us-09-707-919-7

Perfect score: 32
Sequence: 1 cgcacccgcgcctcccgctcgcgcgcgcgt 32

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.0	482	9	AL039573	AL039573 DKF2p34D
2	31	96.9	1100	BM455214	BM455214 AGENCOURT
3	23.6	73.8	673	BF166472	BF166472 601774967
4	23.6	73.8	768	BG489196	BG489196 602518188
5	23	71.9	364	BE457923	BE457923 us99c12.x
6	23	71.9	826	BE547876	BE547876 601074781
7	23	71.9	982	AG071168	AG071168 Pan trogl
8	23	71.9	1201	AG071022	AG071022 Pan trogl
9	22.6	70.6	697	BF864385	BF864385 963051C06
10	22.6	70.6	765	BM051505	BM051505 603638189
11	22.6	70.6	891	AZ186337	AZ186337 SP.1006.B
12	22.6	70.6	1316	AG459258	AG459258 1024024F0
13	22.4	70.0	1285	AG030453	AG030453 Pan trogl
14	22	68.8	440	BE455853	BE455853 SALC_0341
15	22	68.8	570	BM426511	BM426511 pafzn.pk0
16	22	68.8	645	BM426511	BM426511 pafzn.pk0
17	22	68.8	805	AG076652	AG076652 Pan trogl

C 18	22	68.8	842	12	AG126210	AG126210 Pan trogl
C 19	22	68.8	927	9	BE214044	BE214044 HV.CEB000
C 20	22	68.8	984	10	BE573341	BE573341 601333005
C 21	22	68.8	1173	12	AG057265	AG057265 Pan trogl
C 22	22	68.8	1796	10	BE966127	BE966127 601660053
C 23	21.8	68.1	694	10	BE865255	BE865255 963058C05
C 24	21.6	67.5	1000	10	BE702476	BE702476 602684542
C 25	21.4	66.9	328	9	AM428134	AM428134 65010 MAR
C 26	21.4	66.9	641	9	AI881521	AI881521 606070C02
C 27	21.4	66.9	655	10	B1956162	B1956162 HVS.MEM002
C 28	21.4	66.9	704	12	AG063082	AG063082 Pan trogl
C 29	21.4	66.9	847	12	AG076581	AG076581 Pan trogl
C 30	21.4	66.9	892	10	B1182700	B1182700 UNT.-P-FN-
C 31	21.4	66.9	902	9	AL521598	AL521598 AL521598
C 32	21.4	66.9	1002	10	B1730607	B1730607 603351001
C 33	21.4	66.9	1041	10	BF979655	BF979655 602287955
C 34	21.4	66.9	1166	10	BM468435	BM468435 AGENCOURT
C 35	21.4	66.9	1209	12	AG104184	AG104184 Pan trogl
C 36	21.4	66.9	1379	12	AG082668	AG082668 Pan trogl
C 37	21.2	66.2	1201	12	CNS015YM	AL106072 Drosophila
C 38	21	65.6	306	9	BB868471	BB868471 BB868471
C 39	21	65.6	568	9	AI028491	AI028491 OY74C08.x
C 40	21	65.6	801	12	AG058862	AG058862 Pan trogl
C 41	21	65.6	812	9	BE034612	BE034612 ML04A10.M
C 42	21	65.6	842	10	BG339009	BG339009 602436830
C 43	21	65.6	949	12	AG035244	AG035244 Pan trogl
C 44	21	65.6	1007	10	B110052	B110052 602901610
C 45	21	65.6	1121	12	AG078685	AG078685 Pan trogl

ALIGNMENTS

RESULT 1
LOCUS AL039573 482 bp mRNA linear EST 29-FEB-2000
DEFINITION DKF2p34D1311.t1 434 (synonym: htes3) Homo sapiens CDNA clone
ACCESSION AL039573
VERSION DKF2p34D1311.5', mRNA sequence.
KEYWORDS AL039573.1 GI:5408612
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 482)
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
J.S. (Duesterhoeft, et al.)
EST (Duesterhoeft, 1999)
Unpublished (1999)
Contact: Duesterhoeft A
MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKF2p34D1311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.
Location/Qualifiers
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p34D1311"
/clone_1lb="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 49 a 218 c 145 g 70 t

ORIGIN

Query Match 100.0%; Score 32; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgcgcgcgcgt 32
|||||
DB 168 CGCCAAACCCGCCCTCCCGCTCGCGCCGCCG 199

RESULT 2
BM455214 1100 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT 6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
DEFINITION 5' mRNA sequence.
ACCESSION BM455214
VERSION BM455214.1 GI:18504254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1100)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12134 row: k column: 12
High quality sequence stop: 623.

FEATURES
source
1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 240 a 329 c 306 g 219 t 6 others

ORIGIN

Query Match 96.9%; Score 31; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgcgcgcgcgt 31
|||||
DB 142 CGCCAAACCCGCCCTCCCGCTCGCGCCGCCG 172

RESULT 3
BF166472 673 bp mRNA linear EST 30-OCT-2000
LOCUS BF166472/c
DEFINITION 601774967f1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3995513 5',
mRNA sequence.
ACCESSION BF166472
VERSION BF166472.1 GI:11046824
KEYWORDS EST.

SOURCE
house mouse.
Mus musculus

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 673)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9215 row: e column: 18
High quality sequence stop: 672.

FEATURES
source
1..673
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:3995513"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 194 a 119 c 210 g 150 t

ORIGIN

Query Match 73.8%; Score 23.6; DB 10; Length 673;
Best Local Similarity 86.7%; Pred. No. 1.2e+03;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 cgcacccgcgcctcccgctcgcgcgcgcgt 30
|||||
DB 107 CGCCAGCCGCCGCTCTCCCGCTCGCGCCGCC 78

RESULT 4
BG489196 768 bp mRNA linear EST 27-MAR-2001
LOCUS BG489196/c
DEFINITION 602518188f1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4637036 5',
mRNA sequence.
ACCESSION BG489196
VERSION BG489196.1 GI:13450703
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCD/DMP/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1396 row: o column: 21
High quality sequence stop: 631.

SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male BAC library clone:PTB-062B12.R.
ORGANISM	Pan troglodytes
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
TITLE JOURNAL REFERENCE AUTHORS	1 (sites) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. BAC end sequences of Library PTB Unpublished 2 (bases 1 to 982) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. Direct Submission
TITLE JOURNAL REFERENCE AUTHORS	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chumpbes@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/, Tel.:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&d process and may have higher chance of clone tracking errors. PRIMERS
COMMENT	Sequencing: MJ3Rev
FEATURES SOURCE	LIBRARY Vector : pKS145 R.site 1 : SacI R.site 2 : SacI. Location/Qualifiers 1..982 /organism="Pan troglodytes" /db_xref="taxon:9598" /clone="PTB-062B12.R" /sex="male" /cell_type="lymphoblast" /clone_1lb="PTB Chimpanzee Male BAC Library"
BASE COUNT ORIGIN	67 a 243 c 580 g 88 t 4 others
Query Match	71.9%; Score 23; DB 12; Length 982;
Best Local Similarity	83.9%; Pred. No. 1.7e+03;
Matches	26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY	1 cgccaaaccgcgctccccgcgtcgagcggc 31
Db	238 CCCCGAGCCGCCGCCGCCCTCCGC GCCG 208
RESULT	8
LOCUS AGO71022/c	1201 bp DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-062B11.R, genomic survey sequence.
ACCESSION	AGO71022
VERSION	AGO71022.1 GI:16622824
KEYWORDS	GSS; GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male BAC library clone:PTB-062B11.R. Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan. 1 (sites) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. BAC end sequences of Library PTB Unpublished 2 (bases 1 to 1201) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. Direct Submission Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
ORGANISM	Pan troglodytes
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
TITLE JOURNAL REFERENCE AUTHORS	1 (sites) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. BAC end sequences of Library PTB Unpublished 2 (bases 1 to 1201) Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. Direct Submission
TITLE JOURNAL REFERENCE AUTHORS	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

```

COMMENT
(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hnp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PB7. This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.site 1 : SacI
R.site 2 : SacI
Location/Qualifiers
1. 1201
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PB7-062B11.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PB7 Chimpanzee Male BAC library"
BASE COUNT
117 a 251 c 683 g 116 t 34 others
ORIGIN
Query Match 71.9% Score 23; DB 12; Length 1201;
Best Local Similarity 83.9%; Pred. No. 1.7e+03;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 cgcacaccgcgcgcctccgcgcgcgcgcgcgcg 31
||||| ||||| ||||| ||||| |||||
Db 292 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 262

RESULT 9
LOCUS BF864385 697 bp mRNA linear EST 19-JAN-2001
DEFINITION 96305JC06.x1 C. reinhardtii CC-1690, Stress condition I, normalized,
lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF864385
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadales.
1 (bases 1 to 697)
Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C.,
Lefebvre,P., Mcdermott,J.P., Shreger,J., Sillflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
Unpublished (2000)
JOURNAL
Contact: Charles Hauser
COMMENT
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
Location/Qualifiers
1. 697
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, lambda zap II"
/note="Vector: pBluescript II SK-; Site1: EcoRI; Site2:
XhoI; This library, constructed by John Davies and Jeffrey
Mcdermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). POLYA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II

```

(Stratagene) in the EcoRI (5') and XhoI (3') sites. pJuescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldi et al (1996) Genome Research 6: 791-806."

BASE COUNT 165 a 203 c 151 g 178 t
ORIGIN

Query Match 70.6%; Score 22.6; DB 10; Length 697;
Best Local Similarity 86.2%; Pred. No. 2.2e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgcacccgcgcctcccgctgcgcgc 29
||||| ||||||| |||||||
Db 207 CGCCACAGCGCCTCCCGCGCGCGCC 235

RESULT 10 765 bp mRNA linear EST 07-NOV-2001
BM051505/c 603638189F1 NIH_MGC_8 Homo sapiens CDNA clone IMAGE: 419294 5',
LOCUS mRNA sequence.
DEFINITION BM051505
ACCESSION BM051505.1 GI:16780772
VERSION EST.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@bcrfemail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: L1061872 row: a column: 23
High quality sequence stop: 695.

FEATURES
Location/Qualifiers

1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5419294"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 156 a 189 c 257 g 162 t
ORIGIN

Query Match 70.6%; Score 22.6; DB 10; Length 765;
Best Local Similarity 86.2%; Pred. No. 2.2e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 ccaaccgcgcctcccgctgcgcgcgc 31
||||| ||||||| ||||||| |||||||
Db 223 CCACCCCGCGCCACGCGCCTCCCGCGCCG 195

RESULT 11 891 bp DNA linear GSS 30-AUG-2000
A2186337
LOCUS SP.1006.B1.D09.T7A Strongylocentrotus purpuratus, purple sea urchin
DEFINITION 'sperm genomic BAC library Strongylocentrotus purpuratus genomic
clone plate-1006 Col-17 Row-H, DNA sequence.

ACCESSION A2186337.1 GI:8369431
VERSION GSS.
KEYWORDS Strongylocentrotus purpuratus.
SOURCE Strongylocentrotus purpuratus
ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 891)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Blondl,T.R.,
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1006 row: H column: 17
Seq primer: T7
Class: BAC ends
High quality sequence stop: 891.

FEATURES
Location/Qualifiers

1..891
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="plate-1006 Col-17 Row-H"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACs3.6; BAC clones in E-Coli
DH10B"
BASE COUNT 214 a 344 c 193 g 126 t 14 others
ORIGIN

Query Match 70.6%; Score 22.6; DB 12; Length 891;
Best Local Similarity 86.2%; Pred. No. 2.2e+03;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gccacccgcgcctcccgctgcgcgcgc 30
||||| ||||||| ||||||| |||||||
Db 475 GCCACCGCGCGCCCGCGCGCGCCG 503

RESULT 12 1316 bp mRNA linear EST 29-MAY-2001
BG849258
LOCUS 1024024F04.x1 C. reinhardtii CC-1690, normalized, lambda zap II
DEFINITION Chlamydomonas reinhardtii CDNA, mRNA sequence.
BG849258
ACCESSION BG849258.1 GI:14230442
VERSION EST.
KEYWORDS Chlamydomonas reinhardtii.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 1316)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,B., Lefebvre,P.,
McMormott,J.P., Sillflow,C., Stern,D. and Sutzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii genome: A Model,

Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
1..264
/organism="Papio hamadryas"
/db_xref="taxon:9557"
<1..>264
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 25 a 130 c 78 g 31 t
ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 264;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcgcctcccgctcgagcccg 22
|||||
Db 82 GCGCCTCCCGCTCGGCGCCG 103

RESULT 2
ARI59544 355 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59544
DEFINITION Sequence 1 from patent US 6251589.
ACCESSION ARI59544
VERSION ARI59544.1 GI:16222225
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 355)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
JOURNAL Patent: US 6251589-A 1 26-JUN-2001;
FEATURES location/Qualifiers
source 1..355
BASE COUNT 20 a 176 c 102 g 55 t 2 others
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcgcctcccgctcgagcccg 22
|||||
Db 228 GCGCCTCCCGCTCGGCGCCG 249

RESULT 3
AF330028 390 bp DNA linear PRI 08-NOV-2001
LOCUS AF330028
DEFINITION Pan troglodytes SCA2 gene, partial sequence.
ACCESSION AF330028
VERSION AF330028.1 GI:12382830
KEYWORDS chimpanzee.
SOURCE Pan troglodytes
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE 1 (bases 1 to 390)
AUTHORS Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and Brahmachari,S.K.
TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms
JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
PUBMED 11689490
REFERENCE 2 (bases 1 to 390)
AUTHORS Choudhry,S. and Brahmachari,S.K.

TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for Biochemical Technology, Delhi University Campus, Mall Road, Delhi 110 007, India

FEATURES
source
1..390
/organism="Pan troglodytes"
/db_xref="taxon:9598"
repeat_region 1..390
/note="microsatellite"
/rpt_type=tandem
/rpt_unit=cag
<1..>390
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 48 a 183 c 110 g 49 t
ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 390;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcgcctcccgctcgagcccg 22
|||||
Db 76 GCGCCTCCCGCTCGGCGCCG 97

RESULT 4
ARI59558 572 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59558
DEFINITION Sequence 18 from patent US 6251589.
ACCESSION ARI59558
VERSION ARI59558.1 GI:16222251
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 572)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor
JOURNAL Patent: US 6251589-A 18 26-JUN-2001;
FEATURES location/Qualifiers
source 1..572
BASE COUNT 34 a 277 c 174 g 85 t 2 others
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 572;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcgcctcccgctcgagcccg 22
|||||
Db 228 GCGCCTCCCGCTCGGCGCCG 249

RESULT 5
ARI59546 623 bp DNA linear PAT 17-OCT-2001
LOCUS ARI59546
DEFINITION Sequence 5 from patent US 6251589.
ACCESSION ARI59546
VERSION ARI59546.1 GI:16222229
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 623)
AUTHORS Tsuji,S. and Sanpei,K.
TITLE Method for diagnosing spinocerebellar ataxia type 2 and primers therefor

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: US
Center clone name: RP11-42B1

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 224788 bases at least Q40
Consensus quality: 229074 bases at least Q30
Consensus quality: 230948 bases at least Q20
Estimated insert size: 227237; sum-of-coverage estimation
Estimated insert size: 317311; agarose-gel estimation
Estimated coverage: 6.3x in Q20 bases; agarose-gel estimation
Quality coverage: 8.8x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 33241: contig of 33241 bp in length
* 33242 33341: gap of unknown length
* 33342 56391: contig of 23050 bp in length
* 56392 56491: gap of unknown length
* 56492 81323: contig of 24832 bp in length
* 81324 81423: gap of unknown length
* 81424 102538: contig of 21115 bp in length
* 102539 102639: gap of unknown length
* 102639 119710: contig of 17072 bp in length
* 119711 119811: gap of unknown length
* 119811 136913: contig of 17103 bp in length
* 136914 153285: gap of unknown length
* 153286 153385: gap of unknown length
* 153386 167987: contig of 14602 bp in length
* 167988 168087: gap of unknown length
* 168088 178731: contig of 10644 bp in length
* 178732 178831: gap of unknown length
* 178832 186641: contig of 7810 bp in length
* 186642 186741: gap of unknown length
* 186742 193215: contig of 6474 bp in length
* 193216 193315: gap of unknown length
* 193316 201310: contig of 7995 bp in length
* 201311 201410: gap of unknown length
* 201411 208647: contig of 7237 bp in length
* 208648 213802: gap of unknown length
* 213803 213902: contig of 5055 bp in length
* 213903 218049: gap of unknown length
* 218050 218149: gap of unknown length
* 218150 223316: contig of 5167 bp in length
* 223317 223416: gap of unknown length
* 223417 227389: contig of 3973 bp in length
* 227390 227489: gap of unknown length
* 227490 229032: contig of 1543 bp in length
* 229033 229132: gap of unknown length
* 229133 230651: contig of 1519 bp in length
* 230652 231758: gap of unknown length
* 231759 231758: contig of 1007 bp in length.
```

----- Location/Qualifiers

1. 231758
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

Query Match 100.0%; Score 22; DB 2; Length 231758;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggcgtcccccgtcggccgcg 22
|||||
DB 89256 GCGCTCCCGCTCGGCGCCG 89235

RESULT 11
AF330031 303 bp DNA linear PRI 08-NOV-2001

LOCUS Macaca mulatta SCA2 gene, partial sequence.

DEFINITION AF330031

ACCESSION AF330031.1 GI:12382833

VERSION

KEYWORDS

SOURCE

ORGANISM

Thesau monkey.
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE 2 (bases 1 to 322)
 AUTHORS Choudhry, S. and Brahmachari, S.K.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
 Biochemical Technology, Delhi University Campus, Mall Road, Delhi
 110 007, India

FEATURES
 source location/Qualifiers
 1..322
 /organism="Macaca radiata"
 /db_xref="taxon:9548"
 <1..>322
 /gene="SCA2"
 /note="spino cerebellar ataxia 2"

BASE COUNT 32 a 155 c 95 g 40 t
 ORIGIN

Query Match 92.7%; Score 20.4; DB 9; Length 322;
 Best Local Similarity 95.5%; Pred. No. 1.4e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgagcccg 22
 ||||| ||||| ||||| |||||
 Db 105 GCGCCTCCCTGCTCGCGCCCG 126

RESULT 13
 AF330030 384 bp DNA linear PRI 08-NOV-2001
 LOCUS Presbycus entellus SCA2 gene, partial sequence.
 DEFINITION AF330030
 ACCESSION AF330030
 VERSION AF330030.1 GI:12382832
 KEYWORDS Hanuman langur.
 SOURCE Presbytis entellus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Colobinae; Presbytis.
 REFERENCE 1 (bases 1 to 384)
 AUTHORS Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
 Brahmachari, S.K.
 JOURNAL CAG repeat instability at SCA2 locus: anchoring CAA interruptions
 PUMED 11689490
 TITLE Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
 REFERENCE 2 (bases 1 to 384)
 AUTHORS Choudhry, S. and Brahmachari, S.K.
 JOURNAL Direct Submission
 TITLE Submitted (21-DEC-2000) Functional Genomics Unit, Center for
 Biochemical Technology, Delhi University Campus, Mall Road, Delhi
 110 007, India

FEATURES
 source location/Qualifiers
 1..384
 /organism="Presbytis entellus"
 /db_xref="taxon:9574"
 <1..>384
 /gene="SCA2"
 /note="spino cerebellar ataxia 2"

BASE COUNT 46 a 178 c 109 g 51 t
 ORIGIN

Query Match 92.7%; Score 20.4; DB 9; Length 384;
 Best Local Similarity 95.5%; Pred. No. 1.4e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgagcccg 22
 ||||| ||||| ||||| |||||
 Db 82 GCGCCTCCCGCTCAGCGCCCG 103

RESULT 14
 AF330029

LOCUS AF330029 409 bp DNA linear PRI 08-NOV-2001
 DEFINITION Gorilla gorilla SCA2 gene, partial sequence.
 ACCESSION AF330029
 VERSION AF330029.1 GI:12382831
 KEYWORDS
 SOURCE
 ORGANISM Gorilla gorilla
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
 REFERENCE 1 (bases 1 to 409)
 AUTHORS Choudhry, S., Mukerji, M., Srivastava, A.K., Jain, S. and
 Brahmachari, S.K.
 TITLE CAG repeat instability at SCA2 locus: anchoring CAA interruptions
 JOURNAL Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
 PUBMED 11689490
 REFERENCE 2 (bases 1 to 409)
 AUTHORS Choudhry, S. and Brahmachari, S.K.
 JOURNAL Direct Submission
 TITLE Submitted (21-DEC-2000) Functional Genomics Unit, Center for
 Biochemical Technology, Delhi University Campus, Mall Road, Delhi
 110 007, India

FEATURES
 source location/Qualifiers
 1..409
 /organism="Gorilla gorilla"
 /db_xref="taxon:9593"
 <1..>409
 /gene="SCA2"
 /note="spino cerebellar ataxia 2"

BASE COUNT 35 a 196 c 120 g 58 t
 ORIGIN

Query Match 92.7%; Score 20.4; DB 9; Length 409;
 Best Local Similarity 95.5%; Pred. No. 1.4e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgagcccg 22
 ||||| ||||| ||||| |||||
 Db 110 GCGCCTCCCGCTCAGCGCCCG 131

RESULT 15
 SC5E9 29924 bp DNA linear BCT 04-JAN-2001
 LOCUS Streptomyces coelicolor cosmid 5E9.
 DEFINITION A1446003
 ACCESSION A1446003
 VERSION A1446003.1 GI:11061544
 KEYWORDS anti-sigma factor antagonist; DNA-binding; hydrolase; IS110;
 IS1650; Killer toxin-like protein; oxidoreductase; pseudogene;
 secreted; transcriptional regulator; transposase.
 SOURCE Streptomyces coelicolor
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 1 to 29924)
 AUTHORS Redenbach, M., Klier, H.M., Denapate, D., Elchner, A., Cullum, J.,
 Kinsahl, H. and Hopwood, D.A.
 TITLE A set of ordered cosmids and a detailed genetic and physical map
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
 MEDLINE 97000351
 REFERENCE 2 (bases 1 to 29924)
 AUTHORS Seeger, K.J. and Harris, D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 29924)
 AUTHORS Bentley, S.D., Parkhill, J., Barrell, B.G. and Rands, M.A.
 JOURNAL Direct Submission
 TITLE Submitted (26-OCT-2000) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>) CDS are numbered using the following system eg SC787.01c, SC (S. coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of BldB et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 589 lies between and overlaps cosmids 8D11 and 10B8A on the AseI-A genomic restriction fragment.

FEATURES

source

1..29924
/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"

source

1..29924
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"

gene

complement(1..279)
/gene="SC5E9.01c"

misc-feature

1..117
/note="Nominal overlap with Streptomyces coelicolor cosmid 8D11"

CDS

complement(<1..279)
/gene="SC5E9.01c"
/note="SC5E9.01c, unknown, len: 93aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC14481.1"
/db_xref="GI:11061545"

RBS

358..362
/gene="SC5E9.02"

gene

371..754
/gene="SC5E9.02"

CDS

371..754
/gene="SC5E9.02"
/note="SC5E9.02, possible anti-sigma factor antagonist, len: 127aa; weakly similar to many eg. SW:Q9WVX8 (RSW, SRP60) anti-sigma B factor antagonist from Streptomyces coelicolor (113 aa) fasta scores: opt: 118, z-score: 164.6, E(): 0.11, 27.78 identity in 94 aa overlap. Contains Pfam match to entry PF01740 STAS, STAS domain."
/codon_start=1
/transl_table=11
/product="putative anti-sigma factor antagonist"
/protein_id="CAC14482.1"

misc-feature

/db_xref="GI:11061546"
/translation="MSLHKAVTCTAADVSRPGDQASVYLYRCGYPVYNGCEY DLSITPLSGLGTAAREKTKVLEASGITFADSLNLILITONSVDLRVAPARQL RRLIEITGVDAVKRSTVEAATC"

446..742
/gene="SC5E9.02"
/note="Pfam match to entry PF01740 STAS, STAS domain, score 38.20, E-value 1.9e-07"

RBS

901..104
/gene="SC5E9.03"

gene

912..1154
/gene="SC5E9.03"

912..1154
/gene="SC5E9.03"
/note="SC5E9.03, conserved hypothetical protein, len: 80aa; similar to others from Streptomyces coelicolor eg. TR:054206 (EMBL:AJ001206) pepA hypothetical protein from the glycogen metabolism cluster (90 aa) fasta scores: opt: 116, z-score: 182.1, E(): 0.011, 36.3% identity in 80 aa overlap."

RBS

1168..1172
/gene="SC5E9.04"

gene

1175..1399
/gene="SC5E9.04"

1175..1399
/gene="SC5E9.04"
/note="SC5E9.04, unknown, len: 74aa"

CDS

1175..1399
/gene="SC5E9.04"
/note="SC5E9.04, unknown, len: 74aa"

RBS

1175..1399
/gene="SC5E9.04"

gene

1175..1399
/gene="SC5E9.04"

CDS

1175..1399
/gene="SC5E9.04"

misc-feature

complement(2005..2883)
/note="Insertion element IS1650"

gene

complement(2026..2472)
/gene="SC5E9.05c"

CDS

complement(2026..2472)
/gene="SC5E9.05c"

misc-feature

complement(2469..2879)
/gene="SC5E9.05c"

gene

complement(2469..2879)
/gene="SC5E9.05c"

CDS

complement(2469..2879)
/gene="SC5E9.05c"

misc-feature

complement(2469..2879)
/gene="SC5E9.05c"

gene

complement(2469..2879)
/gene="SC5E9.05c"

CDS

complement(2469..2879)
/gene="SC5E9.05c"

misc-feature

complement(2469..2879)
/gene="SC5E9.05c"

```

Query Match
Best Local Similarity 95.5%; Score 20.4; DB 1; Length 29924;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ggcgcctccgcgctcgagccgcg 22
||||| ||||| ||||| |||||
Db 18467 gCGCCTGCCCGCTCGGCGCCG 18488

/misc_feature
/protein_id="CAC14486.1"
/db_xref="GI:11061550"
/translation="MVVGLFVVRHRLTDSMAVIEPLAPRMGRVPDRQVNGI
LMLSTGASMRDLPERYGPKTVYERFRMSADGTDRLLAHVQHSDDAAGVDMTV
CVSTSTRAHQAAGARKRGRTGRASHAGPAAG"
complement(2499..2759)
/gene="SC5E9.06c"
/note="Pfam match to entry PF01511 Transposase_6,
Transposase, score 34.10, E-value 2.3e-07"
2886..2888
/gene="none"
2886..2888
/gene="none"
/note="tla"
/label=""
2968..4563
/gene="SC5E9.07"
2968..4563
/gene="SC5E9.07"
/note="SC5E9.07, possible DNA-binding protein, len: 531aa;
similar to many conserved hypothetical prot. ins eg.
TR:Q9X8M1 (EMBL:AL078610) hypothetical protein from
Streptomyces coelicolor (654 aa) fasta scores: opt: 1459,
z-score: 1666.8, E(): 0.4888 identity in 531 aa overlap.
Contains helix-turn-helix motif (Score 1136 (+3.06 SD)) at

```

Search completed: August 14, 2002, 21:48:55
Job time: 13553 sec

Streptomyces sp. C
Nucleotide sequenc
Human lung tumour
Human lung tumour
Mouse ischaemic co
Human immune/thema
Human lymphocyte
Streptomyces sp. C
Nucleotide sequenc
Nucleotide sequenc
Mycobacterium tube
Human tumour suppl
Human prostate can
Human secreted pro
Wheat Irc1 # 3 cod
Human agrin cDNA,
Human cancer assoc
Drosophila melanog
Human breast cell
Human foetal liver
Probe #449 for ge
Human brain expres
Probe #456 for ge
Probe #4671 used t
Probe #4418 used t
Human prostate tum
Nucleotide sequenc
Product of alterna
Alternatively splic
Human PCDH2 coding
Human T gene DNA f
Human protocadheri
Protocadherin clon
cDNA encoding rat
Skin cell cDNA, SE
DNA encoding novel

CC CAG repeat at a location corresponding to the CAG repeat region of
 CC the SCA2 gene. The presence of at least 13 CAG repeats above the
 CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
 CC Primers (see AAT99640-41) amplifying at least this region are used
 CC for diagnosis. Also claimed are full-length ataxin-2 cDNAs for
 CC human and mouse (see AAV06552-53), kits for detecting mutations at
 CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
 CC useful for studying the physiological roles of SCA2 polypeptide
 CC (ataxin-2, see AAM33807-08) and its effect upon behaviour.

XX Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 516;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgcgcccg 22
 |||
 Db 139 gcgcctcccgctcgcgcccg 160

RESULT 3

AAV17229
 ID AAV17229 standard; DNA; 623 BP.

XX AAV17229;

DT 29-JUN-1998 (first entry)

XX SCA2 gene fragment.

KW SCA2 gene; spinocerebellar ataxia type II; CAG repeat; PCR primer; ss.

XX Synthetic.

FT Key Location/Qualifiers

FT CDS 341..583

FT /tag= a

FT /note= "SCA2 protein fragment, no stop codon given"

PN WO9803679-A1.

PD 29-JAN-1998.

PF 18-JUL-1996; 96WO-JP01999.

PR 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sanpei K, Tsuji S;

PI WPI; 1998-120796/11.

DR P-PSDB; AAM41372.

XX Diagnosing spinocerebellar ataxia type II - by PCR and determining

PT number of CAG repeat units

XX Example 1; Page 11-12; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in

CC the method of the invention for diagnosing spinocerebellar ataxia type

CC II, by performing PCR on the test DNA using two primers hybridising to

CC parts of the SCA2 gene sequence, and determining the number of CAG

CC repeats in the amplified products. The method provides an easy means for

CC the diagnosis of spinocerebellar ataxia type II.

SO Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;

Query Match 100.0%; Score 22; DB 19; Length 623;
 Best Local Similarity 100.0%; Pred. No. 11;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgcgcccg 22
 |||
 Db 228 gcgcctcccgctcgcgcccg 249

RESULT 4

AAT78912
 ID AAT78912 standard; cDNA; 4200 BP.

XX AAT78912;

DT 09-FEB-1998 (first entry)

XX Spinocerebellar ataxia gene SCA2.

KW Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;

KW repeat region; affinity; TATA binding protein; Kennedy disease;

KW transcription initiation factor; lymphoblastic cell line; schizophrenia;

KW Huntington's disease; dominant autosomal spinocerebellar ataxia;

KW X-linked spino-bulbar muscular atrophy; familial spastic paraplegia;

KW dentatorubral-pallidolusial atrophy; bipolar affective disorder;

KW manic depressive psychosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 3..2747

FT /tag= a

FT /product= "SCA2 protein

FT /note= "this CDS contains a putative translational start

FT codon for the SCA2 protein at positions 243-245"

FT CDS 2594..3640

FT /tag= b

FT /note= "this second open reading frame may be derived

FT by a frameshift or by alternative splicing"

FT CDS 239..245

FT /tag= d

FT /note= "putative Kozak consensus signal"

FT repeat_region 258..323

FT /tag= e

FT /note= "encodes polyglutamine repeat region; contains

FT repeats of CAG with 2 CAA codons interspersed"

FT repeat_unit 258..260

FT /tag= f

FT /note= "CAG repeats"

FT misc_feature 1..3986

FT /tag= g

FT /note= "sequence contained in DAN1 clone"

FT misc_feature 3987..4200

FT /tag= h

FT /note= "derived from the EST's AAH92640, AAN90240 and

FT AA213574 from dbEST database"

FT misc_feature 4023..4029

FT /tag= i

FT /note= "region which differs in length between the

FT sequences of the EST clones AAH92640, AAN90240

FT and AA213574"

PN WO9717445-A1.

XX 15-MAY-1997.

PD 08-NOV-1996;

PF 96WO-FR01773.

PR 10-NOV-1995; 95FR-0013576.

PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PI Lutz Y, Mandel J, Tora L, Trotter Y;
 DR MPI: 1997-281034/25.
 DR P-PSDB: AAM24800, AAM24801.
 PT Antibody 1C2 used for treating or preventing neuro-degenerative
 PT diseases - associated with proteins containing long polyglutamine
 PT repeats, e.g. Huntington's disease
 PS
 XX
 XX Claim 21: Page 45-47: 69pp: French.
 CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
 CC treatment of neurodegenerative diseases associated with the presence
 CC of polyglutamine repeat regions. This Mab is already known for its
 CC affinity to the YATA binding protein (YBP) transcription initiation
 CC factor, especially at the amino acid sequence LEEQKQKQKQ found at
 CC the N-terminus of YBP. Mab 1C2 has been shown to have a high affinity
 CC for polyglutamine repeats with a proportional affinity to the number
 CC of glutamine repeats. This affinity has been used to identify genes
 CC encoding proteins containing long polyglutamine repeats which are
 CC implicated in neurodegenerative diseases. A screen of an expression
 CC library, generated from a lymphoblastic cell line from a patient
 CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
 CC new sequences (AAV78906-T78911) encoding polyglutamine repeats. Mab 1C2
 CC also isolated the complete SCA2 gene in clone DAN1 (sequence presented
 CC here). The sequence appears to contain 2 open reading frames (ORF) the
 CC second of which may be generated by an frameshift allpage or by an
 CC alternative splicing event. The first ORF also encodes a 22 amino acid
 CC polyglutamine repeat region near the N-terminus of the protein. Normal
 CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
 CC interspersed whereas the mutant sequence from patients with SCA
 CC contains at least 30, preferably 37-50 CAG repeats.
 CC Mab 1C2, active fragment of it or nucleic acids encoding it are
 CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
 CC X-linked spinobulbar muscular atrophy (Kennedy disease),
 CC dentatorubral-pallidoluysal atrophy, dominant autosomal spinocerebellar
 CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
 CC depressive psychoses and schizophrenia.
 XX
 SO Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

Query Match 100.0%; Score 22; DB 18; Length 4200;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcctcccgctgcgagcccg 22
 ||||||||||||||||||||
 DB 130 gcgcctcccgctgcgagcccg 151

RESULT 5
 AAV30270
 ID AAV30270 standard; DNA: 4367 BP.
 AC AAV30270;
 XX
 DT 02-OCT-1998 (first entry)
 XX
 DE Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
 XX
 KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
 KW CAG repeat; neurodegenerative disease; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 49..3990
 FT /tag= a
 FT /product= "Spinocerebellar ataxia type 2 associated

FT repeat_region 544..612 protein"
 FT FT /tag= b
 FT /note= "normal CAG repeat region; this is increased in
 FT patients with SCA2"
 FT repeat_unit 544..546
 FT FT /tag= c
 FT
 PN MO9818920-A1.
 PD 07-MAY-1998.
 XX
 XX 30-OCT-1997; 97WO-JP03946.
 XX
 XX 30-OCT-1996; 96JP-0304059.
 XX
 XX (SRLS-) SRL INC.
 XX
 XX Sanpei K, Tsuji S;
 XX MPI: 1998-272215/24.
 DR P-PSDB: AAM60213.
 XX
 PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
 PT - contain increased number of CAG repeat region compared to normal
 PT gene
 PS Claim 1; Pages 13-22; 38pp: Japanese.
 XX
 CC This represents the sequence of a gene causative of spinocerebellar
 CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
 CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
 CC expression product produces a polyglutamine sequence from Gln-166 to
 CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
 CC patients this number is increased to 35-100. Peptides encoded by nucleic
 CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
 CC gene, antibodies recognising the peptides and antisense nucleic acids
 CC hybridising with the nucleic acid fragments can be used for the
 CC investigation and diagnosis of SCA2. They can also be used for the
 CC treatment of SCA2 by antisense therapy or gene therapy.
 XX
 SO Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcctcccgctgcgagcccg 22
 ||||||||||||||||||||
 DB 416 gcgcctcccgctgcgagcccg 437

RESULT 6
 AAV06552
 ID AAV06552 standard; cDNA: 4481 BP.
 AC AAV06552;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 DE Human SCA2 cDNA including CAG repeat region.
 XX
 KW SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KW diagnosis; olivo-ponto-cerebellar atrophy; ss; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 164..4101
 FT /tag= a
 FT primer_bind complement (631..648)
 FT /tag= b

```

FT primer_bind /note="primer SCA2-A binding site"
FN 740..757
FT /tag="c
FT /note="primer SCA2-B binding site"
FT primer_bind 1070..1091
FT /tag="d
FT /note="primer SCA2-14B binding site"
FT exon 899..900
FT /tag="e
FT /note="predicted splice site"
FT repeat_region 658..723
FT /tag="f
FT /note="CAG repeat region"
FT repeat_unit 658..660
FT /tag="g
FT /note="CAG repeat"
FT repeat_unit 661..663
FT /tag="h
FT /note="CAG repeat"
FT repeat_unit 664..666
FT /tag="i
FT /note="CAG repeat"
FT repeat_unit 667..669
FT /tag="j
FT /note="CAG repeat"
FT repeat_unit 670..672
FT /tag="k
FT /note="CAG repeat"
FT repeat_unit 673..675
FT /tag="l
FT /note="CAG repeat"
FT repeat_unit 676..678
FT /tag="m
FT /note="CAG repeat"
FT repeat_unit 679..681
FT /tag="n
FT /note="CAG repeat"
FT repeat_unit 685..687
FT /tag="o
FT /note="CAG repeat"
FT repeat_unit 688..690
FT /tag="p
FT /note="CAG repeat"
FT repeat_unit 691..693
FT /tag="q
FT /note="CAG repeat"
FT repeat_unit 694..696
FT /tag="r
FT /note="CAG repeat"
FT repeat_unit 700..702
FT /tag="s
FT /note="CAG repeat"
FT repeat_unit 703..705
FT /tag="t
FT /note="CAG repeat"
FT repeat_unit 706..708
FT /tag="u
FT /note="CAG repeat"
FT repeat_unit 709..711
FT /tag="v
FT /note="CAG repeat"
FT repeat_unit 712..714
FT /tag="w
FT /note="CAG repeat"
FT repeat_unit 715..717
FT /tag="x
FT /note="CAG repeat"
FT repeat_unit 718..720
FT /tag="y
FT /note="CAG repeat"
FT repeat_unit 721..723
FT /tag="z
FT /note="CAG repeat"

```

```

XX MO9742314-A1.
PN 13-NOV-1997.
XX
PD 08-MAY-1997; 97MO-US07725.
XX
PE 08-OCT-1996; 96US-0727084.
XX
PR 08-MAY-1996; 96US-0017388.
PR 19-JUL-1996; 96US-0022207.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Pulst S;
XX
DR WPI: 1998-086523/08.
DR P-PSDB: AAW33807.
XX
PT Nucleic acids encoding human and mouse ataxin 2 - a product of the
PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
XX ataxia type 2
XX
PS Claim 6; Page 52-58; 98pp; English.
XX
CC This cDNA sequence corresponds to a novel SCA2 gene encoding a human
CC spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
CC (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
CC human frontal cortex cDNA library in lambda Zapit were screened
CC with probes obtained by PCR amplification of plasmid AAP512B (see
CC AAW06551). PCR products were used to screen the human adult frontal
CC cortex library, and 5' clones were obtained by RT-PCR of placental
CC mRNAs. Overlapping clones was used to generate the composite 4481
CC bp sequence. Ataxin type 2 can be diagnosed by detecting a genomic
CC or transcribed mRNA sequence in an individual having an expanded
CC CAG repeat at a location corresponding to the CAG repeat region of
CC the SCA2 gene. The presence of at least 13 CAG repeats above the
CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
CC primers (see AAW9640-41) amplifying at least this region are used
CC for diagnosis. Also claimed are kits for detecting mutations at
CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
CC useful for studying the physiological roles of ataxin-2 and its
CC effect upon behaviour.
XX
SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 22; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcctcccgctcgagcccg 22
DB 530 ggcgcctcccgctcgagcccg 551

RESULT 7
AA223428
ID AA223428 standard; DNA; 4481 BP.
XX
AC AA223428;
XX
DT 19-JAN-2000 (first entry)
XX
DE Human SCA2 DNA.
XX
KW Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic; ss.
XX
OS Homo sapiens.

```

```

XX  Key                               Location/Qualifiers
FH  CDS                               163..4101
FT                                     /*tag= a
FT                                     /product= "SCA2"
XX
XX  MO9945944-A1.
XX
XX  16-SEP-1999.
XX
XX  11-MAR-1999; 99WO-US05250.
XX
XX  12-MAR-1998; 98US-0041886.
XX
XX  (BURN-) BURNHAM INST.
XX
XX  Bredezen DE, Rabizadeh S;
XX
XX  WPI; 1999-561617/47.
XX
XX  P-PSDB; AAY33495.
XX
XX  New proapoptotic dependence peptides, used to develop products for
XX  treating, e.g. Alzheimer's disease -
XX
XX  Disclosure: Page 130-135; 199pp; English.
XX
XX  This invention describes novel pure proapoptotic dependence peptides
XX  which comprise a sequence of an active dependence domain selected from
XX  dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
XX  huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
XX  SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
XX  of inducing cell death and can be used to develop products to mediate or
XX  inhibit apoptosis. The methods can be used for reducing the severity of
XX  a proapoptotic dependence domain mediated pathological conditions e.g.
XX  Huntington's disease, Alzheimer's disease, Kennedy's disease,
XX  Spino cerebellar ataxias, dentatorubropallidoluysian atrophy,
XX  Machado-Joseph disease, stroke or head trauma. They can also be used for
XX  reducing the severity of a pathological condition mediated by upregulated
XX  cell proliferation or cell survival e.g. neoplastic, malignant,
XX  autoimmune or fibrotic conditions. This sequence encodes the human
XX  SCA2 polypeptide described in the method of the invention.
XX
XX  Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;
XX
XX  Query Match                               100.0%; Score 22; DB 20; Length 4481;
XX  Best Local Similarity 100.0%; Pred. No. 8.6;
XX  Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 ggcgcctcccgctcgagccgcg 22
XX  ||||||||||||||||||||
XX  Db 530 ggcgcctcccgctcgagccgcg 551
XX
XX  RESULT 8
XX  AAS91762/c
XX  ID AAS91762 standard; CDNA; 588 BP.
XX
XX  AAS91762;
XX
XX  13-FEB-2002 (first entry)
XX
XX  DNA encoding novel human diagnostic protein #27566.
XX
XX  Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX  food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX  Homo sapiens.
XX
XX  WO200175067-A2.
XX
XX  11-OCT-2001.
XX

```

```

PF  30-MAR-2001; 2001WO-US08631.
XX
XX  31-MAR-2000; 2000US-0540217-
XX  23-AUG-2000; 2000US-0649167.
XX
XX  (HYSE-) HYSEQ INC.
XX
XX  Drmanac RT, Liu C, Tang YT;
XX
XX  WPI; 2001-639362/73.
XX
XX  P-PSDB; ABG27575.
XX
XX  New isolated polynucleotide and encoded polypeptides, useful in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits and to assess
XX  biodiversity -
XX
XX  Claim 1; SEQ ID NO 27566; 103pp; English.
XX
XX  The invention relates to isolated polynucleotide (I) and
XX  polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX  and gene mapping, and in recombinant production of (II). The
XX  polynucleotides are also used in diagnostics as expressed sequence tags
XX  for identifying expressed genes. (I) is useful in gene therapy techniques
XX  to restore normal activity of (II) or to treat disease states involving
XX  (II). (II) is useful for generating antibodies against it, detecting or
XX  quantitating a polypeptide in tissue, as molecular weight markers and as
XX  a food supplement. (II) and its binding partners are useful in medical
XX  imaging of sites expressing (II). (I) and (II) are useful for treating
XX  disorders involving aberrant protein expression or biological activity.
XX  The polypeptide and polynucleotide sequences have applications in
XX  diagnostics; forensics; gene mapping; identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. AAS64197-AAS94564 represent novel human
XX  diagnostic coding sequences of the invention.
XX  Note: The sequence data for this patent did not appear in the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX  Sequence 588 BP; 123 A; 175 C; 144 G; 146 T; 0 other;
XX
XX  Query Match                               80.9%; Score 17.8; DB 23; Length 588;
XX  Best Local Similarity 90.5%; Pred. No. 3.8e+02;
XX  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX  2 cgcctcccgctcgagccgcg 22
XX  ||||||||||||||||||
XX  Db 513 CGCCTCCCGCTCGCTCCCGCCG 493
XX
XX  RESULT 9
XX  AAT21643/c
XX  ID AAT21643 standard; CDNA to mRNA; 303 BP.
XX
XX  AAT21643;
XX
XX  06-AUG-1996 (first entry)
XX
XX  Human gene signature H0MG503080.
XX
XX  Gene signature; messenger RNA; mRNA; relative abundance; frequency;
XX  human; cloning; mapping; non-biased library; diagnosis; detection;
XX  cell typing; abnormal cell function; ss.
XX
XX  Homo sapiens.
XX
XX  WO9514772-A1.
XX
XX  01-JUN-1995.
XX

```

PR	08-DEC-1999;	99JP-0348375.
XX	(SETO/) SETO H.	
PA	(KUZU/) KUZUYAMA T.	
XX		
PI	Seto H, Kuzuyama T, Takahashi S, Takagi M;	
XX		
DR	WPI: 2001-381696/40.	
P-PSDB:	AAB99729.	
PT	Actinomycetes-originated genes of enzymes participating in mevalonate pathway, applicable in producing e.g. ubiquinone, vitamin K2 and carotenoids for treatment of heart diseases, osteoporosis and cancer in drug and health food -	
XX		
PS	Claim 4; Page 50-52; 75pp; Japanese.	
XX		
CC	The sequence given in AAH44043 represents a DNA sequence isolated from Streptomyces sp. CL190, containing a 6798 base pairs (S1), which encodes the whole enzyme necessary for functioning the mevalonate pathway. The sequence encodes protein sequences, designated orfA to F and Imgr, which are used in the mevalonate pathway. The proteins and polynucleotide sequences encoding them have cardant, osteopathic and cytostatic activities. The genes are applicable in producing e.g. ubiquinone, vitamin K2 and carotenoids which can be used in the treatment of heart diseases, osteoporosis and cancer in drugs and health foods. The present sequence encodes the orfD protein from the present invention.	
CC		
CC		
CC		
SQ	Sequence 1092 BP; 181 A; 430 C; 340 G; 141 T; 0 other;	
Query Match	78.2%; Score 17.2; DB 22; Length 1092;	
Best Local Similarity	86.4%; Pred. No. 5.7e+02;	
Matches 19; Conservative	0; Mismatches 3; Indels 0; Gaps 0.	
OY	1 gcgcctcccgctcggcgccg 22 	
Db	803 ggccttcgcgcctcggcgccg 824	
RESULT 11		
AAH78257		
ID	AAH78257 standard; DNA; 1104 BP.	
XX		
AC	AAH78257;	
XX		
DT	26-NOV-2001 (first entry)	
XX		
DE	Nucleotide sequence of a Streptomyces sp. CL190 DNA sequence.	
XX		
KM	Non-mevalonate pathway; isopentenyl diphosphate; IPP; antimicrobial; DMAPP; dimethylallyl diphosphate; herbicide; anti-malarial; ss.	
OS	Streptomyces sp.	
XX		
PN	WO200164943-A1.	
XX		
PD	07-SEP-2001.	
XX		
PF	28-FEB-2001: 2001MO-JP01501.	
XX		
PR	02-MAR-2000; 2000JP-0056753.	
XX		
PA	(SETO/) SETO H.	
PA	(KUZU/) KUZUYAMA T.	
XX		
PI	Seto H, Kuzuyama T, Mitsui N;	
XX		
DR	WPI: 2001-570704/64.	
XX		
PT	Screening for substances that inhibit the non-mevalonate pathway comprises using organisms that can use both the mevalonate- and non-mevalonate pathways for synthesizing isopentenyl diphosphate and	

PT dimethylallyl diphosphate -
 XX
 PS Claim 6; Page 36-37; 45pp; Japanese.
 XX
 CC The specification describes a method of screening for substances that
 CC specifically inhibit the non-mevalonate pathway. The method comprises
 CC using organisms that can use both the mevalonate- and non-mevalonate
 CC pathways for synthesizing isopentenyl diphosphate (IPP) and
 CC dimethylallyl diphosphate (DMAPP). The method is used for screening for
 CC substances (e.g., antimicrobials, herbicides and anti-malarials) that
 CC inhibit the non-mevalonate pathway. The present sequence represents
 CC a Streptomyces sp. CLJ90 DNA sequence.
 XX
 SQ Sequence 1104 BP; 182 A; 429 C; 346 G; 147 T; 0 other;
 XX
 Query Match 78.2%; Score 17.2; DB 22; Length 1104;
 Best Local Similarity 86.4%; Pred. No. 5.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcgcctcccgctgcgcgcgcg 22
 DB 815 gcgcctcccgctgcgcgcgcg 836
 XX
 RESULT 12
 AAF68877 standard; cDNA; 2064 BP.
 XX
 AC AAF68877;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:825.
 XX
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytostatic; antisense inhibition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200100828-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 30-JUN-2000; 2000MO-US18061.
 XX
 PR 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MM, Mannion J;
 XX
 DR WPI; 2001-071488/08.
 XX
 PT Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 XX
 PS Claim 4; Page 433-434; 436pp; English.
 XX
 CC The present invention describes immunogenic portions of lung tumour-
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.
 CC (I) have cytostatic activity and can be used in gene therapy, antisense
 CC inhibition and in vaccines. The NAs and the lung tumour-associated

CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with their inappropriate expression,
 CC especially lung cancers. For example, the NAs may be administered to
 CC treat diseases by rectifying mutations or deletions in a patient's genome
 CC or to suppress the activity of the protein by expressing inactive proteins
 CC or to supplement the patient's own production of (I). Additionally, the
 CC NAs may be used to produce the lung-tumour associated protein, according
 CC to standard recombinant DNA methodology. Conversely, antisense NA
 CC molecules may be administered to down regulate protein expression by
 CC binding with the cells own genes and preventing their expression. The NA
 CC and complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar NA sequences in
 CC samples, and hence which patients may be in need of treatment for lung
 CC cancer. The (I) may be used as antigens in the production of antibodies
 CC and in assays to identify modulators (agonists and antagonists) of the
 CC expression and activity of the protein. AAF68878 and AAF68878 and
 CC AAF68848 to AAF68878 represent human lung tumour protein related
 CC nucleotide and protein sequences which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2064 BP; 400 A; 658 C; 627 G; 379 T; 0 other;
 XX
 Query Match 78.2%; Score 17.2; DB 22; Length 2064;
 Best Local Similarity 86.4%; Pred. No. 5.2e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcgcctcccgctgcgcgcgcg 22
 DB 166 gcgcctcccgctgcgcgcgcg 187
 XX
 RESULT 13
 AAF68878 standard; cDNA; 2109 BP.
 XX
 AC AAF68878;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:826.
 XX
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytostatic; antisense inhibition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200100828-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 30-JUN-2000; 2000MO-US18061.
 XX
 PR 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MM, Mannion J;
 XX
 DR WPI; 2001-071488/08.
 XX
 PT Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -

XX Claim 4; Page 434; 436pp; English.
XX The present invention describes immunogenic portions of lung tumour-
XX associated proteins (I) and the nucleic acids (NAs) that encode them.
XX (I) have cytostatic activity and can be used in gene therapy, antisense
XX inhibition and in vaccines. The NAs and the lung tumour-associated
XX proteins they encode may be used in the prevention, treatment and
XX diagnosis of diseases associated with their inappropriate expression,
XX especially lung cancers. For example, the NAs may be administered to
XX treat diseases by rectifying mutations or deletions in a patient's genome
XX that affect the activity of the protein by expressing inactive proteins
XX or to supplement the patient's own production of (I). Additionally, the
XX NAs may be used to produce the lung-tumour associated protein, according
XX to standard recombinant DNA methodology. Conversely, antisense NA
XX molecules may be administered to down regulate protein expression by
XX binding with the cells own genes and preventing their expression. The NA
XX and complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar NA sequences in
XX samples, and hence which patients may be in need of treatment for lung
XX cancer. The (I) may be used as antigens in the production of antibodies
XX and in assays to identify modulators (agonists and antagonists) of the
XX expression and activity of the protein. AAF68083 to AAF68878 and
XX AAB76848 to AAB76878 represent human lung tumour protein related
XX nucleotide and protein sequences which are used in the exemplification
XX of the present invention.
XX Sequence 2109 BP; 424 A; 659 C; 619 G; 407 T; 0 other:
SQ

Query Match 78.2%; Score 17.2; DB 22; Length 2109;
Best Local Similarity 86.4%; Pred. No. 5.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgagccgcg 22
||||| ||| ||| ||| ||| |||
DB 126 gcgcctcccgctcgagccgcg 147

RESULT 14
AB199857/c
ID AB199857 standard; cDNA; 2539 BP.
XX
XX AC AB199857;
XX
XX DT 07-MAR-2002 (first entry)
XX
XX DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:989.
XX
XX KW Mouse; Ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX OS Mus musculus.
XX
XX PN WO200188188-A2.
XX
XX PD 22-NOV-2001.
XX
XX PF 18-MAY-2001; 2001WO-JP04192.
XX
XX PR 18-MAY-2000; 2000JP-0145977.
XX
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
XX DR WPI: 2002-034733/04.
XX
XX DR P-PSDB: ABB57353.
XX
XX PT Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring
XX expression levels of particular genes defined in the specification or
XX PT by determining the expression profile of a gene group comprising these
XX genes -

XX Claim 2; Page 2500-2504; 2690pp; English.
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX genes (I) in a test sample or determining the expression profile of a
XX gene group in the sample comprising genes selected from (I). The method
XX is useful for examining the ischaemic condition (e.g. compressive
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX expression levels of particular genes (AB199202 to AB199917, encoding
XX the protein sequences in ABB57020 to ABB57374) or by determining the
XX expression profile of a gene group comprising these genes. The
XX expression levels or expression profiles produced by these genes are
XX used as an indicator when screening for ischaemic condition-improving
XX drugs or therapeutics for ischaemic diseases. AB199913 and AB199914
XX represent PCR primers for a mouse ischaemic condition related sequence,
XX which are used in the exemplification of the present invention.
XX Sequence 2539 BP; 547 A; 744 C; 770 G; 478 T; 0 other:
SQ

Query Match 78.2%; Score 17.2; DB 24; Length 2539;
Best Local Similarity 86.4%; Pred. No. 5.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcgcctcccgctcgagccgcg 22
||||| ||| ||| ||| ||| |||
DB 201 GGCCTCCCTCTCGGTCGCCG 180

RESULT 15
AAK84365
ID AAK84365 standard; DNA; 5176 BP.
XX
XX AC AAK84365;
XX
XX DT 07-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39177.
XX
XX KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytoskeletal; gene therapy; vaccine; metastasis; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01354.
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX
XX PR 04-FEB-2000; 2000US-0180628.
XX
XX PR 24-FEB-2000; 2000US-0184664.
XX
XX PR 02-MAR-2000; 2000US-0186350.
XX
XX PR 16-MAR-2000; 2000US-0189874.
XX
XX PR 17-MAR-2000; 2000US-0190076.
XX
XX PR 18-APR-2000; 2000US-0198123.
XX
XX PR 19-MAY-2000; 2000US-0205515.
XX
XX PR 07-JUN-2000; 2000US-0209467.
XX
XX PR 28-JUN-2000; 2000US-0214886.
XX
XX PR 30-JUN-2000; 2000US-0215135.
XX
XX PR 07-JUL-2000; 2000US-0216647.
XX
XX PR 07-JUL-2000; 2000US-0216880.
XX
XX PR 11-JUL-2000; 2000US-0217487.
XX
XX PR 11-JUL-2000; 2000US-0217496.
XX
XX PR 14-JUL-2000; 2000US-0218290.
XX
XX PR 26-JUL-2000; 2000US-0220963.
XX
XX PR 26-JUL-2000; 2000US-0220964.
XX
XX PR 14-AUG-2000; 2000US-0224518.
XX
XX PR 14-AUG-2000; 2000US-0224519.
XX
XX PR 14-AUG-2000; 2000US-0225213.
XX
XX PR 14-AUG-2000; 2000US-0225214.

Query Match 78.2%; Score 17.2; DB 22; Length 5176;
 Best Local Similarity 86.4%; Pred. No. 4.6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 gcgcctcccgctcgcccg 22
 ||||||| |||||||
 db 2938 gctctccgctcgcccg 2959

Search completed: August 14, 2002, 22:06:48
 Job time: 11703 sec


```

: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 623
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (341)..(583)
: FEATURE:
: OTHER INFORMATION: Tsp-2
: OS-09-043-303-5

```

Query Match	100.0%;	Score 22;	DB 4;	Length 623;
Best Local Similarity	100.0%;	Pred. No. 1.8;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 gcgcctcccgctcggcccg 22
|||
Db 228 gcgcctcccgctcggcccg 249

RESULT 3
US-09-041-886-18
; Sequence 18, Application US/09041886

```

1  GENERAL INFORMATION:
2
3  APPLICANT: Bredesen, Dale E.
4
5  APPLICANT: Rabizadeh, Sharoz
6
7  TITLE OF INVENTION: Proapoptotic Peptides, Dependence
8
9  TITLE OF INVENTION: Polypeptides and Methods of Use
10
11 NUMBER OF SEQUENCES: 72

```

ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 7000
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 26268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:

```

? SEQUENCE CHARACTERISTICS:
?     LENGTH: 4481 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?     MOLECULE TYPE: DNA (genomic)
?     FEATURE:
?     NAME/KEY: CDS
?     LOCATION: 163..4099
US-09-041-886-18

```

Query Match	100.0%	Score 22	DB 4	Length 4481
Best Local Similarity	100.0%	Pred. No. 1.5		
Matches 22; Conservative	0;	Mismatches	0;	Indels 0;
Gaps	0;			

Db 530 GCGCCTCCCCGCTCGGCGCCCG 551

RESULT 4
US-09-103-840A-2/C

```

; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

```

```

1  TITLE OF INVENTION:  DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
2  TITLE OF INVENTION:  TUBERCULOSIS
3  FILE REFERENCE:  24366-20007 00
4  CURRENT APPLICATION NUMBER:  US/09/103,840A
5  CURRENT FILING DATE:  1998-06-24

```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
```

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

```

OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match	78.28;	Score 17.2;	DB 4;	Length 4403765;
Best Local Similarity	86.48;	Pred. No. 27;		
Matches 19; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY 1 gcgcctcccgctcgcgccg 22
 ||||| | |||| | |||||
 Db 437224 GCGCGCGCGCGCGCGCGCG 437203

RESULT 5
US-08-644-271-31
; Sequence 31, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

ADDRESSEE: Regeneron Pharmaceuticals, Inc.
ADDRESS: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette
 ? COMPUTER: IBM compatible
 ? OPERATING SYSTEM: DOS
 ? SOFTWARE: FastSO Version 2.0
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/644,271
 ? FILING DATE: 10-MAY-1996
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:

Db 530 GCGCCTCCCCGCTCGGCGCCCG 551

RESULT 4
US-09-103-840A-2/C

```

; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

```

```

1  TITLE OF INVENTION:  DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
2
3  TITLE OF INVENTION:  TUBERCULOSIS
4
5  FILE REFERENCE:  24366-20007  00
6
7  CURRENT APPLICATION NUMBER:  US/09/103,840A
8
9  CURRENT FILING DATE:  1998-06-24

```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
```

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

```

OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match	78.28;	Score 17.2;	DB 4;	Length 4403765;
Best Local Similarity	86.48;	Pred. No. 27;		
Matches 19; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY 1 gcgcctcccgctcgagcccg 22
||||| 1 ||||| ||||| |||||
Db 437224 GCGCGCGCGCGCGCGCGCGCG 437203

RESULT 5
US-08-644-271-31
; Sequence 31, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

ADDRESSEE: Regeneron Pharmaceuticals, Inc.
ADDRESS: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette
 ? COMPUTER: IBM compatible
 ? OPERATING SYSTEM: DOS
 ? SOFTWARE: FastSO Version 2.0
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/644,271
 ? FILING DATE: 10-MAY-1996
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:

TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1479 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..1476
OTHER INFORMATION:
NAME/KEY: Human Agt1n
LOCATION: 1..1479
OTHER INFORMATION:
US-08-644-271-31

Query Match 76.4%; Score 16.8; DB 1; Length 1479;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcctcccgctcgagccg 20
11111111111111111111
DB 1268 GCCTCCCGCGCGCGCC 1287

RESULT 6
US-07-998-003A-104/C
Sequence 104, Application US/07998003A
Patent No. 5643781
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
STREET: 20 South Clark Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/998,003A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5643781and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 30903
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/346-5750
TELEFAX: 312/984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-07-998-003A-104

Query Match 76.4%; Score 16.8; DB 1; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ggcctcccgctcgagccg 22
11111111111111111111
DB 2361 GCCTCCCGCGCGCGCCG 2342

RESULT 7
US-08-453-274B-104/C
Sequence 104, Application US/08453274B
Patent No. 5663300
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,274B
FILING DATE: 30-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5663300and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32660
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-08-453-274B-104
Query Match 76.4%; Score 16.8; DB 1; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ggcctcccgctcgagccg 22
11111111111111111111
DB 2361 GCCTCCCGCGCGCGCCG 2342

RESULT 8
US-08-453-695A-104/C
Sequence 104, Application US/08453695A
Patent No. 5708143
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESSEE: Borun
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,695A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: No. 5708143and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32658
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-08-453-695A-104

Query Match 76.4%; Score 16.8; DB 1; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gccctcccgctcgagcccg 22
DB 2361 GCCTCCCGCAGCGCTCCG 2342

RESULT 9
US-08-268-161A-104/C
Sequence 104, Application US/08268161A
Patent No. 5798224
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESS: Borun
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/268,161A
FILING DATE: June 27, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Young J. Suh

REGISTRATION NUMBER: P-41,337
REFERENCE/DOCKET NUMBER: 27866/32149
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-08-268-161A-104

Query Match 76.4%; Score 16.8; DB 1; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gccctcccgctcgagcccg 22
DB 2361 GCCTCCCGCAGCGCTCCG 2342

RESULT 10
US-08-453-702A-104/C
Sequence 104, Application US/08453702A
Patent No. 5891706
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESS: Borun
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,702A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5891706and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32657
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-08-453-702A-104

Query Match 76.4%; Score 16.8; DB 2; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gctcccgctcgagcccg 22
|||||
Db 2361 GCCTCCCGCAGCGCTCCG 2342

RESULT 11

PCT-US93-12588-104/C
Sequence 104, Application US/09099639
Patent No. 626237
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
STREET: 233 South Wacker, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,639
FILING DATE: 18 JUN 1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/263,161
FILING DATE: 27 JUN 1994
ATTORNEY/AGENT INFORMATION:
NAME: Greta E. No. 626237 and
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/34703
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-09-099-639-104

Query Match 76.4%; Score 16.8; DB 4; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gctcccgctcgagcccg 22
|||||
Db 2361 GCCTCCCGCAGCGCTCCG 2342

RESULT 12
PCT-US93-12588-104/C
Sequence 104, Application PC/TUS9312588
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/12588
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/998,003
FILING DATE: 29 DEC 1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31811
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
PCT-US93-12588-104

Query Match 76.4%; Score 16.8; DB 5; Length 2789;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gctcccgctcgagcccg 22
|||||
Db 2361 GCCTCCCGCAGCGCTCCG 2342

RESULT 13

PCT-US95-08071-104/C
Sequence 104, Application PC/TUS9508071
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08071

```

;
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12588
; FILING DATE: 23 DEC 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/998,003
; FILING DATE: 29 DEC 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32149
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ. ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2789 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..2622
; PCT-US95-08071-104

Query Match 76.4%: Score 16.8; DB 5; Length 2789;
Best Local Similarity 90.0%: Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gccctcccgctcgccgccg 22
|||||
DB 2361 GCCTCCCGCAGCGGCTCG 2342

RESULT 14
US-07-998-003A-96/c
; Sequence 96, Application US/07998003A
; Patent No. 5643781
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; Bicknell
; STREET: 20 South Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/998,003A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5643781and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 30903
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ. ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4705 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..2827
; US-08-453-274B-96

Query Match 76.4%: Score 16.8; DB 1; Length 4705;
Best Local Similarity 90.0%: Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gccctcccgctcgccgccg 22
|||||
DB 2361 GCCTCCCGCAGCGGCTCG 2342
```

```

;
; LENGTH: 4705 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..2827
; US-07-998-003A-96

Query Match 76.4%: Score 16.8; DB 1; Length 4705;
Best Local Similarity 90.0%: Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gccctcccgctcgccgccg 22
|||||
DB 2361 GCCTCCCGCAGCGGCTCG 2342

RESULT 15
US-08-453-274B-96/c
; Sequence 96, Application US/08453274B
; Patent No. 5663300
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,274B
; FILING DATE: 30-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5663300and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32660
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ. ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4705 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..2827
; US-08-453-274B-96

Query Match 76.4%: Score 16.8; DB 1; Length 4705;
Best Local Similarity 90.0%: Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gccctcccgctcgccgccg 22
|||||
DB 2361 GCCTCCCGCAGCGGCTCG 2342
```

Thu Aug 15 09:03:16 2002

us-09-707-919-8.rni

Page 7

Search completed: August 14, 2002, 21:57:31
Job time: 13884 sec

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcctcccgctcgagccgcg 22
|||||
Db 177 GCGCTCCCGCTCGCGCCG 198

RESULT 2
BM455214 1100 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
DEFINITION 5', mRNA sequence.
ACCESSION BM455214 GI:18504254
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 1100)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12134 row: k column: 12
High quality sequence stop: 623.
Location/Qualifiers
1.1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_1ib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: PCMV-SPOK6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 240 a 329 c 306 g 219 t 6 others

ORIGIN

Query Match 100.0%; Score 22; DB 10; Length 1100;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcctcccgctcgagccgcg 22
|||||
Db 151 GCGCTCCCGCTCGCGCCG 172

RESULT 3
BE281531 665 bp mRNA linear EST 13-JUL-2000
LOCUS 601155125F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3138342 5',
DEFINITION mRNA sequence.
ACCESSION BE281531
VERSION BE281531.1 GI:9156552
KEYWORDS
EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 665)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1C0M103 row: b column: 07
High quality sequence stop: 585.
Location/Qualifiers
1.665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3138342"
/clone_1ib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pORF7; Site:1; XhoI;
Site:2; EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCGACGAC(G). Size-selected
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 130 a 170 c 255 g 110 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 665;
Best Local Similarity 90.9%; Pred. No. 4.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcctcccgctcgagccgcg 22
|||||
Db 220 GCGCTCCCGCTCGCGCCG 199

RESULT 4
BF865255 694 bp mRNA linear EST 19-JAN-2001
LOCUS 963058C05.y1 C. reinhardtii CC-1690, Stress condition I, normalized
DEFINITION , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF865255
VERSION BF865255.1 GI:12255399
KEYWORDS
SOURCE EST.
ORGANISM Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shrago, J., Sillflow, C., and Stern, D.
Analyses of the Chlamydomonas reinhardtii genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

AG044395/c 1013 bp DNA linear GSS 02-NOV-2001
 LOCUS AG044395
 DEFINITION Pan troglodytes DNA, clone: PTB-022P24.F, genomic survey sequence.
 ACCESSION AG044395
 VERSION AG044395.1 GI:16581212
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone:PTB-022P24.F.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 REFERENCE 1 (sites)
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1013)
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1013)
 DIRECT SUBMISSION
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB this BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
 PRIMERS
 Sequencing: -21M13
 LIBRARY Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..1013
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-022P24.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"
 262 c 321 g 119 t 17 others
 BASE COUNT 294 a
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 1013;
 Best Local Similarity 90.9%; Pred. No. 4.4e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB this BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
 PRIMERS
 Sequencing: M13Rev
 LIBRARY Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..1024
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-021C12.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"
 418 c 233 g 179 t 86 others
 BASE COUNT 108 a
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 1024;
 Best Local Similarity 90.9%; Pred. No. 4.4e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

0Y 1 ggcgcctccgcctcgagcccg 22
 Db 93 GGGCCCCCGCGCGCGCCG 114

RESULT 9 255 bp mRNA linear EST 05-APR-1999
 A1578916
 LOCUS A1578916
 DEFINITION UT-R-G0-ut-c-12-0-UT-62 UT-R-G0 Rattus norvegicus cDNA clone
 UT-R-G0-ut-c-12-0-UT 3', mRNA sequence.
 A1578916
 ACCESSION A1578916
 VERSION A1578916.1 GI:4563292
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 255)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence is likely internal to the message. cDNA library Preparation: M.B. Soares lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 Location/Qualifiers
 1..255
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"

RESULT	11
LOCUS	AV670674/c
DEFINITION	AV670674 OLNI cell line cDNA library (OLB) Oryzias latipes cDNA clone OLB6.03c similar to cathepsin D (EC 3.4.23.5) precursor (chicken), mRNA sequence.
ACCESSION	AV670674
VERSION	AV670674.1 GI:9936472
KEYWORDS	EST.
SOURCE	Japanese medaka.
ORGANISM	Oryzias latipes
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percormorpha; Athrinomorpha; Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
AUTHORS	Naruse,K., Tanaka,M., Shima,A. and Mitani,H.
TITLE	Naruse,K., Tanaka,M., Shima,A. and Mitani,H.
JOURNAL	Medaka EST Project in University of Tokyo
COMMENT	Unpublished (2000)
FEATURES	Contact: Kiyoshi Naruse Department of Biological Sciences Graduate School of Science, University of Tokyo Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan Tel: 81-3-5841-4443 Fax: 81-3-5841-4410 Email: naruse@biol.s.u-tokyo.ac.jp This clone was isolated from OLNI cell line cDNA library (OLB) end sequences. Location/Qualifiers 1..371 /organism="Oryzias latipes" /strain="HN1" /db_xref="taxon:8090" /clone="OLB6.03c" /clone_1lb="OLNI cell line cDNA library (OLB)" BASE COUNT 71 a 128 c 101 g 71 t ORIGIN
QUERY MATCH	Query Match 83.6%; Score 18.4; DB 9; Length 371; Best Local Similarity 95.0%; Pred. No. 5.7e+03; Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB	1 gcgcctcccgctcgagccc 20 11 Db 60 gcgcctcccgctcgagccc 41
RESULT	12
LOCUS	BJ005251
DEFINITION	BJ005251 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA076H03 5'

ACCESSION BJ005251
 VERSION BJ005251.1 GI:17358406
 KEYWORDS EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 414)
 Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
 Medaka EST Project in Takeda's lab
 Unpublished (2001)
 Contact: Tadasu Shin-1
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..414
 /organism="Oryzias latipes"
 /strain="Hd-R"
 /db_xref="taxon:8090"
 /clone_1db="MF01SSA076H03"
 /clone_1id="MF01SSA CDNA"
 /sex="mixture of female and male"
 /tissue_type="whole embryo"
 /dev_stage="segmentation stage 20 - 25"
 BASE COUNT 85 a 143 c 112 g 74 t
 ORIGIN

Query Match 83.6%; Score 18.4; DB 10; Length 414;
 Best Local Similarity 95.0%; Pred. No. 5.7e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ggcctcccgctgcgcgc 20
 ||||||||||||||||
 Db 71 GCGCTCCCGCTGCAGCC 52

RESULT 13
 A1706594 455 bp mRNA linear EST 03-JUN-1999
 LOCUS A1706594
 DEFINITION UI-R-AE1-zf-c-12-0-UI-s1 UI-R-AE1 Rattus norvegicus CDNA clone
 UI-R-AE1-zf-c-12-0-UI 3', mRNA sequence.
 ACCESSION A1706594
 VERSION A1706594.1 GI:4994494
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 455)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@iuii.uiowa.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. CDNA library preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive

elements were found in this CDNA sequence: 289-374, >(CGG
)nSimple repeat
 Seq primer: M13 Forward
 POLYA-No.
 FEATURES
 source Location/Qualifiers
 1..455
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AE1-zf-c-12-0-UI"
 /clone_1db="UI-R-AE1"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AE1
 library is a normalized library constructed from 15 dpc
 rat ventricle. The tag is a string of 5 nucleotides
 present between the Not I site and the oligo-dT track.
 The library was constructed as described by Bonaldo,
 Lennon and Soares, Genome Research 6: 791-806, 1996.
 Tissue provided by Jim Lin, Department of Biology,
 University of Iowa.
 TAG_SEQ=None found"
 BASE COUNT 63 a 194 c 145 g 53 t
 ORIGIN

Query Match 83.6%; Score 18.4; DB 9; Length 455;
 Best Local Similarity 95.0%; Pred. No. 5.7e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 cgcctcccgctgcgcgc 21
 ||||||||||||||||
 Db 227 GCGCTCCCGCTGCAGCC 246

RESULT 14
 BJ006483/c 463 bp mRNA linear EST 05-DEC-2001
 LOCUS BJ006483
 DEFINITION BJ006483 MF01SSA CDNA Oryzias latipes CDNA clone MF01SSA094E08 5',
 mRNA sequence.
 ACCESSION BJ006483
 VERSION BJ006483.1 GI:17362643
 KEYWORDS EST.
 SOURCE Japanese medaka.
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 463)
 Kohara,Y., Shin-I,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
 Medaka EST Project in Takeda's lab
 Unpublished (2001)
 Contact: Tadasu Shin-1
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..463
 /organism="Oryzias latipes"
 /strain="Hd-R"
 /db_xref="taxon:8090"
 /clone="MF01SSA094E08"
 /clone_1db="MF01SSA CDNA"
 /sex="mixture of female and male"
 /tissue_type="whole embryo"
 /dev_stage="segmentation stage 20 - 25"
 BASE COUNT 92 a 159 c 122 g 90 t
 ORIGIN

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:48:55 ; Search time 2563.92 Seconds
(without alignments)
220.372 Million cell updates/sec

Title: US-09-707-919-9

Sequence: 1 ccccttcgtctctctcttccccct 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

```
Minimum DB seq length: 0
```

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Maximum Match 100%

Listing first 45 summaries

Database :

1:	gb_emb1.*
2:	gb_ntg.*
3:	gb_in.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pat.*
7:	gb_ph.*
8:	gb_pl.*
9:	gb_pro.*
10:	gb_ro.*
11:	gb_sts.*
12:	gb_sy.*
13:	gb_un.*
14:	gb_vl.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mu.*
20:	em_om.*
21:	em_or.*
22:	em_ov.*
23:	em_pat.*
24:	em_ph.*
25:	em_pl.*
26:	em_ro.*
27:	em_sts.*
28:	em_un.*
29:	em_vl.*
30:	em_hum.*
31:	em_ntg_inv.*
32:	em_htg_other.*
33:	em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB	ID	Description
---------------	----------------	-----------------	----	----	-------------

C	45	22.2	82.2	2191913	2	AC074307	Mus muscu
C	44	22.2	82.2	210816	2	AC027813	Hom sapi
C	43	22.2	82.2	203957	9	AP003227	Oryza sat
C	42	22.2	82.2	199503	2	AC018901	Hom sapi
C	41	22.2	82.2	195893	2	AC106576	Rattus no
C	35	22.2	82.2	167031	2	AC090233	Hom sapi
C	36	22.2	82.2	167556	2	AC104419	Hom sapi
C	37	22.2	82.2	171571	2	AC097690	Rattus no
C	38	22.2	82.2	183304	2	AC087874	Mus muscu
C	39	22.2	82.2	188246	2	AC013565	Hom sapi
C	40	22.2	82.2	193892	9	AC073037	Hom sapi
C	34	22.2	82.2	195893	2	AC018901	Hom sapi
C	33	22.2	82.2	167018	2	AC073317	Hom sapi
C	32	22.2	82.2	165893	2	AC106576	Rattus no
C	33	22.2	82.2	165893	2	AC090233	Hom sapi
C	35	22.2	82.2	167031	2	AC104419	Hom sapi
C	36	22.2	82.2	167556	2	AC104419	Hom sapi
C	37	22.2	82.2	171571	2	AC097690	Rattus no
C	38	22.2	82.2	183304	2	AC087874	Mus muscu
C	39	22.2	82.2	188246	2	AC013565	Hom sapi
C	40	22.2	82.2	193892	9	AC073037	Hom sapi
C	41	22.2	82.2	195893	2	AC018901	Hom sapi
C	42	22.2	82.2	203957	9	AP003227	Oryza sat
C	43	22.2	82.2	210816	2	AC013564	Hom sapi
C	44	22.2	82.2	2191913	2	AC027813	Hom sapi
C	45	22.2	82.2	220469	2	AC074307	Mus muscu
C	1	27	100.0	264	9	AF330032	Papio ham
C	2	27	100.0	384	9	AF330030	Presbytis
C	3	27	100.0	390	9	AF330028	Pan trogl
C	4	27	100.0	409	9	AF330029	AF330029
C	5	27	100.0	231758	2	AC004085	Gorilla g
C	6	26	96.3	303	9	AF330031	Hom sapi
C	7	26	96.3	322	9	AF330033	Macaca mu
C	8	25.4	94.1	4163	6	HSDNCSA2	Macaca ra
C	9	25.4	94.1	4200	6	AS2706	H.sapiens m
C	10	25.4	94.1	4481	6	AR153580	Sequence 7
C	11	25.4	94.1	4481	6	HSU70323	Human ataxi
C	12	25	92.6	335	6	AR159544	Sequence
C	13	25	92.6	572	6	AR159558	Sequence
C	14	25	92.6	623	6	AR159546	Sequence
C	15	22.2	82.2	2233	9	AK056159	Hom sapi
C	16	22.2	82.2	3908	9	AK056477	Hom sapi
C	17	22.2	82.2	24238	9	AC053546	Hom sapi
C	18	22.2	82.2	62129	2	AC110235	Mus muscu
C	19	22.2	82.2	85444	2	AP003900	Hom sapi
C	20	22.2	82.2	90130	2	AL161618	Human DNA
C	21	22.2	82.2	110000	2	LMFLCHR2..08	Contribution (9 of
C	22	22.2	82.2	112180	2	HSJ32263	AL096677 Human DNA
C	23	22.2	82.2	114067	2	AC093974	Rattus no
C	24	22.2	82.2	118396	9	AC073138	Hom sapi
C	25	22.2	82.2	128484	9	AC025270	Hom sapi
C	26	22.2	82.2	137205	2	AP003261	Oryza sat
C	27	22.2	82.2	138070	2	AC109783	Mus muscu
C	28	22.2	82.2	151441	30	AC036158	Hom sapi
C	29	22.2	82.2	157017	9	AL162389	Human DNA
C	30	22.2	82.2	157371	2	AC106392	Rattus no
C	31	22.2	82.2	157665	9	AC104695	Hom sapi
C	32	22.2	82.2	162018	2	AC092610	Hom sapi
C	33	22.2	82.2	162701	2	AC073317	Hom

ALIGNMENTS

RESULT	1
AF330032	
LOCUS	264 bp DNA linear PRI 08-NOV-2001
DEFINITION	Papio hamadryas SCA2 gene, partial sequence.
ACCESSION	AF330032
VERSION	AF330032.1 GI:12382934
KEYWORDS	
SOURCE	baboon.
ORGANISM	Papio hamadryas

REFERENCE AUTHORS	1 (bases 1 to 264) Choudhury,S., Mukerji,M., Srivastava,A.K., Jain,S. and Brahmachari,S.K.
TITLE	CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms
JOURNAL PUBMED	Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
REFERENCE AUTHORS	2 (bases 1 to 264) Choudhury,S. and Brahmachari,S.K.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-2000) Functional Genomics Unit, Center for

Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
location/Qualifiers
1..264
/organism="Papio hamadryas"
/db_xref="taxon:9557"
<1..>264
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT
25 a 130 c 78 g 31 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 27; DB 9; Length 264;
Pred. No. 5.4;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgctctctctccccc 27
|||||
Db 20 CCCCTTCGTCTCTCTCTCCCT 46

RESULT 2
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS Presbytlis entellus SCA2 gene, partial sequence.
DEFINITION AF330030
ACCESSION AF330030.1 GI:12382832
VERSION
KEYWORDS
SOURCE Hanuman langur.
ORGANISM Presbytlis entellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Colobinae; Presbytis.
1 (bases 1 to 384)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

JOURNAL
PUBMED 11689490
REFERENCE 2 (bases 1 to 384)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
location/Qualifiers
1..384
/organism="Presbytis entellus"
/db_xref="taxon:9574"
<1..>384
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT
46 a 178 c 109 g 51 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 27; DB 9; Length 384;
Pred. No. 5.2;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgctctctctccccc 27
|||||
Db 20 CCCCTTCGTCTCTCTCTCCCT 46

RESULT 3
AF330028 390 bp DNA linear PRI 08-NOV-2001
LOCUS Pan troglodytes SCA2 gene, partial sequence.
DEFINITION AF330028
ACCESSION AF330028.1 GI:12382830
VERSION

KEYWORDS
SOURCE chimpanzee.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 390)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

JOURNAL
PUBMED 11689490
REFERENCE 2 (bases 1 to 390)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
location/Qualifiers
1..390
/organism="Pan troglodytes"
/db_xref="taxon:9598"
1..390
/note="microsatellite"
/rpt_type="tandem
/rpt_unit="cag
<1..>390
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT
48 a 183 c 110 g 49 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 27; DB 9; Length 390;
Pred. No. 5.2;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgctctctctccccc 27
|||||
Db 14 CCCCTTCGTCTCTCTCTCCCT 40

RESULT 4
AF330029 409 bp DNA linear PRI 08-NOV-2001
LOCUS Gorilla gorilla SCA2 gene, partial sequence.
DEFINITION AF330029
ACCESSION AF330029.1 GI:12382831
VERSION
KEYWORDS
SOURCE
ORGANISM gorilla.
Gorilla gorilla
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
1 (bases 1 to 409)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)

JOURNAL
PUBMED 11689490
REFERENCE 2 (bases 1 to 409)
AUTHORS Choudhry,S. and Brahmachari,S.K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India

FEATURES
source
location/Qualifiers
1..409
/organism="Gorilla gorilla"
/db_xref="taxon:9593"
<1..>409
/gene="SCA2"
/note="spinocerebellar ataxia 2"

BASE COUNT 35 a 196 c 120 g 58 t
ORIGIN

Query Match 100.0% Score 27: DB 9: Length 409:
Best Local Similarity 100.0% Pred. No. 5.2:
Matches 27: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 ccccttcgctgcctcctccctc 27
|||||
DB 48 ccccttcgctgcctcctcctc 74

RESULT 5
AC004085/c
LOCUS
DEFINITION AC004085 231758 bp DNA linear HTG 06-NOV-2000
Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered
pieces

AC004085 GI:11079383
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 231758)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
Alspbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbraia,J.,
Benton,J., Blincoe,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Delhorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.R., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunartine,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsil,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jacksen,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudh,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korva,J.,
Koval,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lie,C., Liu,J., Liu,W.,
Loulisege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., Moleod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabhat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokankwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quilis,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshchari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wellington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Morley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 231758)
Submitted (30-JAN-1998)

Direct Submission
Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 3, 2000 this sequence version replaced gi:966929.

FEATURES
source

BASE COUNT 64974 a 51086 c 51148 g 62641 t 1909 others

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: UG
Center clone name: RP11-42B1

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224788 bases at least Q40
Consensus quality: 229074 bases at least Q30
Consensus quality: 230948 bases at least Q20
Estimated insert size: 227237; sum-of-contigs estimation
Estimated insert size: 317311; agarose-gel estimation
Quality coverage: 6.3x in Q20 bases; agarose-gel estimation
Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 33241: contig of 33241 bp in length
33242 33441: gap of unknown length
33442 56391: contig of 23050 bp in length
56392 56491: gap of unknown length
56492 81323: contig of 24832 bp in length
81324 81423: gap of unknown length
81424 102538: contig of 21115 bp in length
102539 102638: gap of unknown length
102639 119710: contig of 17072 bp in length
119711 119810: gap of unknown length
119811 136913: contig of 17103 bp in length
136914 153285: gap of unknown length
153286 153385: contig of 16772 bp in length
153386 153385: gap of unknown length
153386 167987: contig of 14602 bp in length
167988 168087: gap of unknown length
168088 178731: contig of 10644 bp in length
178732 178831: gap of unknown length
178832 186641: contig of 7810 bp in length
186642 186741: gap of unknown length
186742 193215: gap of 6474 bp in length
193216 193315: gap of unknown length
193316 201310: contig of 7995 bp in length
201311 208647: gap of unknown length
208648 208747: gap of unknown length
208748 213802: contig of 5055 bp in length
213803 213902: gap of unknown length
213903 218049: contig of 4147 bp in length
218050 218149: gap of unknown length
218150 223316: contig of 5167 bp in length
223317 223416: gap of unknown length
223417 227389: contig of 3973 bp in length
227390 227490: gap of unknown length
227490 229032: contig of 1543 bp in length
229033 229132: gap of unknown length
229133 230651: contig of 1519 bp in length
230652 230751: gap of unknown length
230752 231758: contig of 1007 bp in length.

Location/Qualifiers
1. 231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"

AUTHORS	Bredesen, D. E. and Razibadeh, S.
TITLE	Proapoptotic peptides dependence polypeptides and methods of use
JOURNAL	Patent: US 6235872-A 18-22-MAY-2001.
FEATURES	Location/Qualifiers
Source	1..4481
BASE COUNT	1144 a 1380 c 1014 g 943 t
ORIGIN	
Query Match	94.1%; Score 25.4; DB 6; Length 4481;
Best Local Similarity	96.3%; Pred. No.15;
Matches	26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 cccctcgtcgtccctccttcctccccc 27
Db	468 CCCCTGCTGCTGCTCCTTCCCCCT 494
RESULT 11	
LOCUS	HSU70323 4481 bp mRNA linear PRI 20-NOV-1996
DEFINITION	Human ataxin-2 (SCA2) mRNA, complete cds.
ACCESSION	U70323
VERSION	U70323.1 GI:16796683
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 4481)
TITLE	Pulst, S.-M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.-N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., DeJong, P., Rouleau, G. A., Auburger, G., Kornberg, J. R., Figueroa, C. and Sabba, S.
MEDLINE	Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2
REFERENCE	Nature Genet. 14 (3), 269-276 (1996)
TITLE	2 (bases 1 to 4481)
JOURNAL	Pulst, S.-M.
AUTHORS	Direct Submission
JOURNAL	Submitted (10-SEP-1996) Medicine, Cedars-Sinai, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
FEATURES	Location/Qualifiers
Source	1..4481
gene	/organism="Homo sapiens"
gene	/db_xref="taxon:9606"
gene	/chromosome="12"
gene	/map="12q24.1"
gene	1..4481
gene	/gene="SCA2"
gene	163..4101
gene	/gene="SCA2"
gene	/standard_name="spinocerebellar ataxia type 2"
gene	/codon_start=1
gene	/product="ataxin-2"
gene	/protein_id="AAB19200.1"
gene	/db_xref="GI:16796684"
gene	/translation="MRSAAAPRSVAVATESRRFAAANWPGRIQAPARSGRGCGG
gene	AAPGYPSPAAPPPGPPSPKQSSPSPASDCFGNGNGGGAFRGSRRLUGCPPI
gene	PAFVVLILASPGAPPAARTASPLIGARSPRGVSLARAPGCCPRACEPVYDPL
gene	MSLFRQOQOQOQOQOQOQOQOQOQOQPPAANVRKPGGGLASPAAPSSSS
gene	SSSSATAPSSSVVATGGGPGGLGRGNKGLPGSTISFDGIYANMVAHILFVVA
gene	SKCEYVONGGIYEGVFRTYSPKCLVIDAAHEKSTESSGGRKEIMELIFKCSDI
gene	VYVQPKMDSSIAKRDAPFDSAISAKVNGEHEKDELPMWDELINAELELVNDV
gene	NGMDNDMFRINENTGVSTIDSSLSSTVYLERDENSEEFLKRRANQLAELETES
gene	AOYKARVALENDRSEEEKYTAIVQRNNSEREGHSINTENKTIIPGQRREYISWS
gene	RONSRRMGQSGSGMPSRSTSHTSDFENPNSGSDQVYNGVWPSPCSPSPSR
gene	QSGPSLPAPRATPTRPSPRSPRSPRSPSHSPASGSPAPVSTMKRMSSEGPPMAS
gene	KAORIKRHRVAVAGSGISSGLEFVSHNPSPAAATPAARTASPGSGTASVGVPR
gene	SPKTRPRSPRONSIGNTPSGVPLASPOAGIIPTPEAVAMPYPAASPPTASVSNRAV
gene	PSSEAKDSRLDQRONSPAGNKENIKPNETSPSPSKAENKGISPVVSEHRKIDDLK

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'-3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source
Location/Qualifiers
1..2233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2R12002541"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2R12"
/note="cloning vector: pME18FLJ-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"
15..1757
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71107.1"
/db_xref="GI:16551486"
/translation="MLRSVLELQNFSESOEYEVSSQVSNAPAPRASVATDGPAPS
CUTDSLDPGERASNGMSDLPPKSKDKLDKKKFKVPPYPIKIRASGRLAGRKVFV
ETPKKRTTRRLREQOKTAEQDGYRCPODSDPRVTGEMEPVSKNGCOAGAELEL
SKAGPEEEDEEEDERDEGEKKSNFKCSICEKALYEKSFIAKSKHRRGVATEVYTR
CDTCGQTFANRCNLKSHQHVHSSERHFPCELCGKFKKKDKVKRHLQVHEGGERH
RCGCGKGLSSKTLALHRTHTGDPYGCCTECGARFSOPSAIKTMR1HTGKRPVC
DECGARFTONHMLIYHKRCHTGERPFCETCGSFASKEXYLKHNRIHTGSKPFCFV
CPRTPAORNSLYOHIKVHTGERPYCCDOCGKQFTOLNALORHRIHTGERPFCNACG
PRTPTKSTLRHTSHDKNTPKKSFIVYDSSPKNDGKHTEQDPEYVSSKLSDKL
SFAENGHFHNLAAYVDYPTMQENSSADTACKADSVYSQDTLATITSELSELTPTQ
DSMPYQLHSLSMNE"

BASE COUNT 610 a 552 c 592 g 479 t
ORIGIN

Query Match 82.2%; Score 22.2; DB 9; Length 2233;
Best Local Similarity 88.9%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttcgtctctccttcctccct 27
|||||
DB 552 CCCCTTCTTGCTCTCTCTCTCT 526

Search completed: August 14, 2002, 21:49:05
Job time: 13563 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 22:06:48 ; Search time 906.46 Seconds

(without alignments)
51.140 Million cell updates/sec

Title: US-09-707-919-9

Perfect score: 27

Sequence: 1 cccctcgtcgtcctcctcctccct 27

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.4	94.1	516	19	SCA2 gene fragment
2	25.4	94.1	4200	18	Spinocerebellar at
3	25.4	94.1	4367	19	gene causative of
4	25.4	94.1	4481	19	Human SCA2 CDNA in
5	25.4	94.1	4481	20	Human SCA2 CDNA. H
6	25	92.6	355	19	SCA2 gene fragment
7	25	92.6	623	19	SCA2 gene fragment
8	22.2	82.2	21724	22	Human genomic DNA
9	22.2	82.2	21724	22	Human Immune/naema

C	10	22.2	82.2	21727	22	AA526630	Human genomic DNA
C	11	22.2	82.2	21727	22	AAK66126	Human Immune/naema
C	12	21.2	78.5	405	21	AAK31280	Plant microsatell
C	13	21.2	78.5	465	22	ABA43043	Human breast cell
C	14	21.2	78.5	465	22	ABA54558	Human foetal liver
C	15	21.2	78.5	465	22	ABA23228	Probe #1694 for ge
C	16	21.2	78.5	465	22	AAK01727	Human brain expres
C	17	21.2	78.5	465	22	AAK27181	Human bone marrow
C	18	21.2	78.5	465	22	AA111767	Probe #1700 for ge
C	19	21.2	78.5	465	22	AA133078	Probe #1764 used t
C	20	21.2	78.5	465	22	AA101696	Probe #1687 used t
C	21	21.2	78.5	1711	21	AA191911	Human ORFX ORF2695
C	22	21.2	78.5	53552	22	AA513655	Genomic DNA sequen
C	23	20.6	76.3	131	21	AAK12116	Human secreted pro
C	24	20.6	76.3	166	22	ABA70491	Human foetal liver
C	25	20.6	76.3	166	22	ABA71579	Human foetal liver
C	26	20.6	76.3	166	22	ABA37121	Probe #15587 for g
C	27	20.6	76.3	166	22	AAK18736	Human brain expres
C	28	20.6	76.3	166	22	AAK19911	Human brain expres
C	29	20.6	76.3	166	22	AAK44670	Human bone marrow
C	30	20.6	76.3	166	22	AAK45945	Human bone marrow
C	31	20.6	76.3	166	22	AA124963	Probe #14896 for g
C	32	20.6	76.3	166	22	AA150650	Probe #19336 used
C	33	20.6	76.3	166	22	AA151859	Probe #20545 used
C	34	20.6	76.3	207	22	ABA48460	Human breast cell
C	35	20.6	76.3	207	22	ABA66363	Human foetal liver
C	36	20.6	76.3	207	22	ABA33425	Probe #11891 for g
C	37	20.6	76.3	207	22	AAK14780	Human brain expres
C	38	20.6	76.3	207	22	AAK40521	Human bone marrow
C	39	20.6	76.3	207	22	AA121280	Probe #11213 for g
C	40	20.6	76.3	207	22	AA146555	Probe #15241 used
C	41	20.6	76.3	207	22	AA106986	Probe #6977 used t
C	42	20.6	76.3	300	20	AA213093	Human gene express
C	43	20.6	76.3	419	23	AA580334	DNA encoding novel
C	44	20.6	76.3	465	23	ABL26183	Drosophila melanog
C	45	20.6	76.3	466	22	ABA57881	Human foetal liver

ALIGNMENTS

RESULT 1

ID AAV06551 standard; DNA: 516 BP.

AC AAV06551:

DT 06-JUL-1998 (first entry)

XX SCA2 gene fragment including CAG repeat region.

DE

XX SCA2 gene: spinocerebellar ataxia-2; ataxin-2; human;

KW diagnosis; olivo-ponto-cerebellar atrophy; ss; ds.

KW

XX Homo sapiens.

OS

XX Key

FH primer_bind

FT location/Qualifiers

FT complement (241..257)

FT primer_bind

FT /tag- a

FT /note- "primer SCA2-A binding site"

FT primer_bind

FT /tag- b

FT /note- "primer SCA2-B binding site"

FT exon

FT /tag- c

FT /note- "predicted splice site"

FT repeat_region

FT /tag- d

FT /note- "CAG repeat region"

FT repeat_unit

FT /tag- e

FT /note- "CAG repeat"

FT repeat_unit

FT 270..272

PT	Nucleic acids encoding human and mouse ataxin 2 - a product of the									
PT	spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of									
PT	ataxia type 2									
PS	Example 2: Page 51-52; 98pp; English.									
XX										
XX										
CC	This genomic DNA in plasmid pL65122B includes a CAG repeat region									
CC	from the novel human SCA2 gene (see AAV06552). It was identified									
CC	following the construction of a bacterial artificial chromosome									
CC	contig and a pl artificial chromosome of the spinocerebellar									
CC	ataxia 2 (SCA2) gene region and the identification of the SCA2									
CC	gene from this contiguous map unit using a technique that screens									
CC	for the presence of DNA trinucleotide repeats. The SCA2 locus is									
CC	at 12q24.1. Ataxia type 2 can be diagnosed by detecting a genomic									
CC	or transcribed mRNA sequence in an individual having an expanded									
CC	CAG repeat at a location corresponding to the CAG repeat region of									
CC	the SCA2 gene. The presence of at least 13 CAG repeats above the									
CC	normal level (22, occasionally 23, repeats) is indicative of SCA2.									
CC	Primers (see AAG99640-41) amplifying at least this region are used									
CC	for diagnosis. Also claimed are full-length ataxin-2 cDNAs for									
CC	human and mouse (see AAV06552-53), kits for detecting mutations at									
CC	the SCA2 locus, antisense oligonucleotides, and transgenic animals									
CC	useful for studying the physiological roles of SCA2 polypeptide									
CC	(ataxin-2, see AAW33807-08) and its effect upon behaviour.									
XX										
XX										
SQL	Sequence 516 BP; 50 A; 228 C; 166 G; 72 T; 0 other;									
QY										
Query Match	94.1%;	Score 25.4;	DB 19;	Length 516;						
Best Local Similarity	96.3%;	Pred. No. 3.1;								
Matches 26; Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;			
1 ccccttcgctgcctctctctccccc 27										
77 ccccttcgctgcctctctctccccc 103										
RESULT 2										
AA78912										
ID	AA78912 standard; CDNA; 4200 BP.									
XX										
XX										
AC	AA78912;									
XX										
DT	09-FEB-1998 (first entry)									
XX										
DE	Spinocerebellar ataxia gene SCA2.									
XX										
KW	Monoclonal antibody; neurodegenerative disease; polyglutamine; TBP;									
KW	repeat region; affinity; TARA binding protein; Kennedy disease;									
KW	transcription initiation factor; lymphoblastic cell line; schizophr									
KW	Huntington's disease; dominant autosomal spinocerebellar ataxia;									
KW	X-linked spino-bulbar muscular atrophy; familial spastic paraplegia;									
KW	dentatorubral pallidolusial atrophy; bipolar affective disorder;									
KW	manic depressive psychosis; ss.									
XX										
OS	Homo sapiens.									
XX										
XX										
PH	Key									
FT	CDS									
FT	Location/Qualifiers									
FT	3..2747									
FT	/tag- a									
FT	/product- SCA2 protein									
FT	/note- "this CDS contains a putative translational start									
FT	codon for the SCA2 protein at positions 243-245"									
FT	CDS									
FT	2594..3640									
FT	/tag- b									
FT	/note- "this second open reading frame may be derived									
FT	by a frameshift or by alternative splicing"									
FT	CDS									
FT	3..242									
FT	/tag- c									
FT	/note- "putative open reading frame which is in frame									
FT	with the putative translational start site of									
FT	the SCA2 open reading frame"									
FT	misc_signal									
FT	239..245									


```

FT      /*tag= d
FT      /note= "putative Kozak consensus signal"
FT      258..323
FT      /*tag= e
FT      /note= "encodes polyglutamine repeat region; contains
FT      repeats of CAG with 2 CAA codons interspersed"
FT      repeat_unit
FT      258..260
FT      /*tag= f
FT      /note= "CAG repeats"
FT      1..3986
FT      /*tag= g
FT      /note= "sequence contained in DAN1 clone"
FT      3987..4200
FT      /*tag= h
FT      /note= "derived from the EST's AAN92640, AAN90240 and
FT      AAZ13574 from dbEST database"
FT      misc_feature
FT      4023..4029
FT      /*tag= i
FT      /note= "region which differs in length between the
FT      sequences of the EST clones AAN92640, AAN90240
FT      and AAZ13574"
FT      misc_feature
FT      4023..4029
FT      /*tag= i
FT      /note= "region which differs in length between the
FT      sequences of the EST clones AAN92640, AAN90240
FT      and AAZ13574"
FT      MO9717445-A1.
FT      15-MAY-1997.
FT      PD
FT      08-NOV-1996; 96WO-FR01773.
FT      PF
FT      10-NOV-1995; 95FR-0013576.
FT      PR
FT      (CNRS ) CNRS CENT NAT RECH SCI.
FT      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
FT      PA
FT      Lutz Y, Mandel J, Tora L, Trollier Y;
FT      PI
FT      WPI: 1997-281034/25.
FT      DR
FT      P-PSDB; AAM24800, AAM24801.
FT      PT
FT      Antibody 1C2 used for treating or preventing neuro-degenerative
FT      diseases - associated with proteins containing long poly:glutamine
FT      repeats, e.g. Huntington's disease
FT      PS
FT      Claim 21; Page 45-47; 69pp; French.
XX      CC
XX      CC The invention relates to a monoclonal antibody (Mab) 1C2 for the
XX      CC treatment of neurodegenerative diseases associated with the presence
XX      CC of polyglutamine repeat regions. This Mab is already known for its
XX      CC affinity to the TARA binding protein (TBP) transcription initiation
XX      CC factor, especially at the amino acid sequence LEEQQRQ000Q found at
XX      CC the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
XX      CC for polyglutamine repeats with a proportional affinity to the number
XX      CC of glutamine repeats. This affinity has been used to identify genes
XX      CC encoding proteins containing long polyglutamine repeats which are
XX      CC implicated in neurodegenerative diseases. A screen of an expression
XX      CC library, generated from a lymphoblastic cell line from a patient
XX      CC suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
XX      CC new sequences (AA78906-778911) encoding polyglutamine repeats. Mab 1C2
XX      CC also isolated the complete SCA2 gene in clone DAN1 (sequence presented
XX      CC here). The sequence appears to contain 2 open reading frames (ORF) the
XX      CC second of which may be generated by a frameshift slippage or by an
XX      CC alternative splicing event. The first ORF also encodes a 22 amino acid
XX      CC polyglutamine repeat region near the N-terminus of the protein. Normal
XX      CC SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
XX      CC interspersed whereas the mutant sequence from patients with SCA
XX      CC contains at least 30, preferably 37-50 CAG repeats.
XX      CC Mab 1C2, active fragment of it or nucleic acids encoding it are
XX      CC specifically used to treat Huntington's disease, SCA types 1-5 or 7,
XX      CC x-linked spino-bulbar muscular atrophy (Kennedy disease),
XX      CC dentatorubral-pallidoluysial atrophy, dominant autosomal spinocerebellar
XX      CC ataxia, familial spastic paraplegia, bipolar affective disorder, manic
XX      CC depressive psychoses and schizophrenia.
XX      CC
XX      Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

```

```

Query Match          94.1%; Score 25.4; DB 18; Length 4200;
Best Local Similarity 96.3%; Pred. No. 3.1;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      1 cccctgcgtccctccctccccc 27
        ||||||||||| |||||||||||
Db       68 cccctgcgtccctccctccccc 94

RESULT 3
AAV30270
ID      AAV30270 standard; DNA; 4367 BP.
XX      AC
XX      AAV30270;
XX      02-OCT-1998 (first entry)
XX      DE
XX      Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
XX      KM
XX      Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
XX      KM
XX      CAG repeat; neurodegenerative disease; ds.
XX      OS
XX      Homo sapiens.
XX      FH
XX      Key      Location/Qualifiers
XX      CDS      49..3990
XX      FT      /*tag= a
XX      FT      /product= "Spinocerebellar ataxia type 2 associated
XX      FT      repeat_region 544..612
XX      FT      /*tag= b
XX      FT      /note= "normal CAG repeat region; this is increased in
XX      FT      repeat_unit 544..546
XX      FT      /*tag= c
XX      PN
XX      MO9818920-A1.
XX      PD
XX      07-MAY-1998.
XX      PF
XX      30-OCT-1997; 97WO-JP03946.
XX      PR
XX      30-OCT-1996; 96JP-0304059.
XX      PA
XX      (SRLS-) SRL INC.
XX      PI
XX      Sanpei K, Tsuji S;
XX      WPI: 1998-272215/24.
XX      DR
XX      P-PSDB; AAM60213.
XX      PT
XX      Nucleic acid fragments associated with spinocerebellar ataxia type 2
XX      gene
XX      - contain increased number of CAG repeat region compared to normal
XX      gene
XX      Claim 1; Pages 13-22; 38pp; Japanese.
XX      CC
XX      CC This represents the sequence of a gene causative of spinocerebellar
XX      CC ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
XX      CC with SCA2, has a tri-nucleotide (CAG) repeat region which in the
XX      CC expression product produces a polyglutamine sequence from Gln-166 to
XX      CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
XX      CC patients this number is increased to 35-100. Peptides encoded by nucleic
XX      CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
XX      CC gene; antibodies recognising the peptides and antisense nucleic acids
XX      CC hybridising with the nucleic acid fragments can be used for the
XX      CC investigation and diagnosis of SCA2. They can also be used for the
XX      CC treatment of SCA2 by antisense therapy or gene therapy.
XX      CC
XX      Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

```

```

Query Match      94.1%; Score 25.4; DB 19; Length 4367;
Best Local Similarity 96.3%; Fred. No. 3.1;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ccccttcgctgcctcctcctccct 27
Db 354 ccccttcgctgcctcctcctccct 380

RESULT 4
AAV06552
ID AAV06552 standard; cDNA; 4481 BP.
XX
AC AAV06552;
XX
DT 06-JUL-1998 (first entry)
XX
XX Human SCA2 cDNA including CAG repeat region.
XX
XX SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
KM diagnosis; olivoponto-cerebellar atrophy; ss; ds.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FH CDS 164..4101
FT /tag= a
FT primer_bind complement (631..648)
FT /tag= b
FT /note= "primer SCA2-A binding site"
FT 740..757
FT /tag= c
FT /note= "primer SCA2-B binding site"
FT 1070..1091
FT /tag= d
FT /note= "primer SCA2-14B binding site"
FT 899..900
FT /tag= e
FT /note= "predicted splice site"
FT 658..723
FT /tag= f
FT /note= "CAG repeat region"
FT 658..660
FT /tag= g
FT /note= "CAG repeat"
FT 661..663
FT /tag= h
FT /note= "CAG repeat"
FT 664..666
FT /tag= i
FT /note= "CAG repeat"
FT 667..669
FT /tag= j
FT /note= "CAG repeat"
FT 670..672
FT /tag= k
FT /note= "CAG repeat"
FT 673..675
FT /tag= l
FT /note= "CAG repeat"
FT 676..678
FT /tag= m
FT /note= "CAG repeat"
FT 679..681
FT /tag= n
FT /note= "CAG repeat"
FT 685..687
FT /tag= o
FT /note= "CAG repeat"
FT 688..690
FT /tag= p
FT /note= "CAG repeat"

```

```

FT repeat_unit 691..693
FT /tag= q
FT /note= "CAG repeat"
FT 694..696
FT /tag= r
FT /note= "CAG repeat"
FT 700..702
FT /tag= s
FT /note= "CAG repeat"
FT 703..705
FT /tag= t
FT /note= "CAG repeat"
FT 706..708
FT /tag= u
FT /note= "CAG repeat"
FT 709..711
FT /tag= v
FT /note= "CAG repeat"
FT 712..714
FT /tag= w
FT /note= "CAG repeat"
FT 715..717
FT /tag= x
FT /note= "CAG repeat"
FT 718..720
FT /tag= y
FT /note= "CAG repeat"
FT 721..723
FT /tag= z
FT /note= "CAG repeat"

WO9742314-A1.
PD 13-NOV-1997.
XX
PF 08-MAY-1997; 97WO-US07725.
PR 08-OCT-1996; 96US-0727084.
PR 08-MAY-1996; 96US-0017388.
PR 19-JUL-1996; 96US-0022207.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Pulat S;
XX
PI MPI: 1998-086523/08.
XX
DR P-PSDB; AAW33807.
XX
XX Nucleic acids encoding human and mouse ataxin 2 - a product of the
PT spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of
PT ataxia type 2
XX
PS Claim 6: Page 52-58; 98pp; English.
XX
XX This cDNA sequence corresponds to a novel SCA2 gene encoding a human
CC spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2
CC (see AAW33807). A trisomy 21 foetal brain cDNA library and an adult
CC human frontal cortex cDNA library in lambda Zapri were screened
CC with probes obtained by PCR amplification of plasmid AAW512B (see
CC AAV06551). PCR products were used to screen the human adult frontal
CC cortex library, and 5' clones were obtained by RT-PCR of placental
CC mRNAs. Overlapping clones was used to generate the composite 4481
CC bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic
CC or transcribed mRNA sequence in an individual having an expanded
CC CAG repeat at a location corresponding to the CAG repeat region of
CC the SCA2 gene. The presence of at least 13 CAG repeats above the
CC normal level (22, occasionally 23, repeats) is indicative of SCA2.
CC primers (see AAT99640-41) amplifying at least this region are used
CC for diagnosis. Also claimed are kits for detecting mutations at
CC the SCA2 locus, antisense oligonucleotides, and transgenic animals
CC useful for studying the physiological roles of ataxin-2 and its
CC effect upon behaviour.
XX

```

SO Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 94.1%; Score 25.4; DB 19; Length 4481;
Best Local Similarity 96.3%; Pred. No. 3.1;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccccttcgtcgtcctcctcctccct 27
|||||
Db 468 ccccttcgtcgtcctcctcctccct 494

RESULT 5

AA223428
ID AA223428 standard; DNA; 4481 BP.

XX AA223428;

DT 19-JAN-2000 (first entry)

XX Human SCA2 DNA.

XX Prosopoptic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCAB;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 163..4101

FT /*tag= a

FT /product= "SCA2"

XX WO9945944-A1.

XX 16-SEP-1999.

XX 11-MAR-1999; 99WO-US05250.

XX 12-MAR-1998; 98US-0041886.

XX (BURN-) BURNHAM INST.

XX Bredeesen DE, Rabizadeh S;

XX WPI; 1999-561617/47.

XX P-PSDB; AAY33495.

XX New prosopoptic dependence peptides, used to develop products for

XX treating, e.g. Alzheimer's disease -

XX Disclosure: Page 130-135; 1999p: English.

XX This invention describes novel pure prosopoptic dependence peptides
CC which comprise a sequence of an active dependence domain selected from
CC dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC SCAB and atrophin-1 polypeptide. The prosopoptic peptides are capable
CC of inducing cell death and can be used to develop products to mediate or
CC inhibit apoptosis. The methods can be used for reducing the severity of
CC a prosopoptic dependence domain mediated pathological conditions e.g.
CC Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC Spinocerebellar ataxias, dentatorubropallidoluysian atrophy,

CC Machado-Joseph disease, stroke or head trauma. They can also be used for
CC reducing the severity of a pathological condition mediated by upregulated
CC cell proliferation or cell survival e.g. neoplastic, malignant,
CC autoimmune or fibrotic conditions. This sequence encodes the human
CC SCA2 polypeptide described in the method of the invention.

XX Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 94.1%; Score 25.4; DB 20; Length 4481;
Best Local Similarity 96.3%; Pred. No. 3.1;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccccttcgtcgtcctcctcctccct 27
|||||
Db 468 ccccttcgtcgtcctcctcctccct 494

RESULT 6

AAV17224
ID AAV17224 standard; DNA; 355 BP.

XX AAV17224;

DT 29-JUN-1998 (first entry)

XX SCA2 gene fragment.

XX SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 341..355

FT /*tag= a

FT /note= "SCA2 protein fragment"

XX WO9803679-A1.

XX 29-JAN-1998.

XX 18-JUL-1996; 96WO-JP01999.

XX 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sanpei K, Tsuji S;

XX WPI; 1998-120796/11.

XX P-PSDB; AAN41370.

XX Diagnosing spinocerebellar ataxis type II - by PCR and determining

XX number of CAG repeat units

XX Claim 1; Page 10; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in

XX the method of the invention for diagnosing spinocerebellar ataxis type

XX II, by performing PCR on the test DNA using two primers hybridizing to

XX parts of the SCA2 gene sequence, and determining the number of CAG

XX repeats in the amplified products. The method provides an easy means for

XX the diagnosis of spinocerebellar ataxis type II.

XX Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

XX Query Match 92.6%; Score 25; DB 19; Length 355;

XX Best Local Similarity 92.6%; Pred. No. 4.3;

XX Matches 25; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

XX OY 1 ccccttcgtcgtcctcctcctccct 27

XX |||||

XX Db 166 ccccttcgtcgtcgtcctcctccct 192

XX RESULT 7

XX AAV17229

XX ID AAV17229 standard; DNA; 623 BP.

XX XX

AC AAV17229;
XX
XX 29-JUN-1998 (first entry)
XX
XX SCA2 gene fragment.
DE
XX SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX
XX Synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT CDS 341..383
FT /tag=a
FT /note="SCA2 protein fragment, no stop codon given"
XX
XX W09803679-A1.
XX
XX 29-JAN-1998.
XX
XX 18-JUL-1996; 96WO-JP01999.
XX
XX 18-JUL-1996; 96WO-JP01999.
XX
XX (SRLS-) SRL INC.
XX
XX Sanpei K, Tsuji S;
XX
XX WPI: 1998-120796/11.
DR P-PSDB; AAW41372.
XX
XX
XX Diagnosing spinocerebellar ataxis type II - by PCR and determining
PT number of CAG repeat units
XX
XX Example 1; Page 11-12; 23pp; Japanese.
XX
XX This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxis type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxis type II.
XX
XX Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;
SO

Query Match 92.6%; Score 25; DB 19; Length 623;
Best Local Similarity 92.6%; Pred. No. 4.3;
Matches 25; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccccttcgtctctctctctccct 27
|||||
DB 166 ccccttcgtctctctctccct 192

RESULT 8
AAS26629/C
ID AAS26629 standard; DNA: 21724 BP.
XX
XX AAS26629;
AC
XX
XX 07-NOV-2001 (first entry)
XX
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1603.
XX
XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;
XX cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnery; secreted protein; rheumatoid arthritis;
XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX corneal infection; wound healing; epithelial cell proliferation;
XX skin ageing; food additive; preservative; antiproliferative.

XX
XX Homo sapiens.
XX
XX W0200155322-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01341.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226881.
XX 22-AUG-2000; 2000US-0226886.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234497.
XX 25-SEP-2000; 2000US-0234498.
XX 25-SEP-2000; 2000US-0234498.

PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246619.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI, 2001-488783/53.
 XX
 DR
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Disclosure; SEQ ID NO 1603; 980bp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence is a genomic DNA encoding a partial novel secreted protein of
 CC the invention.
 CC
 Query Match 82.2%; Score 22.2; DB 22; Length 21724;
 Best Local Similarity 88.9%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 cccctcgtcgtcctcctcctccccc 27
 Db 374 CCCCTCTGCTCCTCCTCCTCCTCCT 348
 ID AAK86125 standard; DNA; 21724 BP.
 RESULT 9
 AAK86125/c
 AC AAK86125;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40937.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR XX

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226281.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231142.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0248207.
PR 17-NOV-2000; 2000US-0248208.
PR 17-NOV-2000; 2000US-0248209.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure: SEQ ID NO 40937; 3071pp + Sequence Listing: English.
CC
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cystostatic

CC activity, and can be used in gene therapy and vaccine production. (1)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (1) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (1) by expressing inactive proteins or to
CC supplement the patient's own production of (1). Additionally, (1)
CC polynucleotides may be used to produce the secreted (1), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 21724 BP; 6980 A; 4481 C; 4432 G; 5831 T; 0 other;

Query Match 82.2%; Score 22.2; DB 22; Length 21724;
Best Local Similarity 88.9%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ccccttggtgctctctctcccccct 27
||||||| ||||||| ||||| |||
Db 374 CCCCTCTCGTCCTCCTCCTCCT 348

RESULT 10
AAS26630
ID AAS26630 standard; DNA; 21727 BP.

XX AAS26630;

DT 07-NOV-2001 (first entry)

DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1604.

XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;
XX cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnerrary; secreted protein; rheumatoid arthritis;
XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX corneal infection; wound healing; epithelial cell proliferation;
XX skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.

XX WO200155322-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01341.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216880.

XX 07-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 11-JUL-2000; 2000US-0218290.

XX 14-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237049.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246613.
PR	08-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249269.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251855.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254059.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-488783/53.	
XX		
PT	New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives -	
PT		
PS	Disclosure: SEQ ID No 1604; 980bp; English.	
XX		
XX	The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi	
CC		

[illegible]


```

RESULT 12
AAA31280
ID AAA31280 standard; DNA: 405 BP.
XX
XX
AC AAA31280;
XX
XX 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #241.
XX
XX Plant microsatellite sequence; core repeat sequence; detection; probe;
XX DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX variety identification; genetic variability evaluation; primer; ss.
XX
XX Eucalyptus grandis.
XX
XX WO9967421-A1.
XX
XX 29-DEC-1999.
XX
XX 25-JUN-1999; 99WO-N200092.
XX
XX 25-JUN-1998; 98US-0105307.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Havukkala IJ, Bloksberg LN, Glenn M;
XX
XX WPI: 2000-116958/10.
XX
XX New plant microsatellite markers and associated flanking species for
XX the detection of polymorphic genetic markers -
XX
XX Claim 1; Page 146-147; 392pp; English.
XX
XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
XX and associated flanking species. The sequences comprise a central core
XX repeat sequence, especially selected from the sequences AAA31094-A32096
XX with left and right flanking sequences. The polynucleotide sequences
XX can be used in the detection of DNA polymorphisms, in genome mapping,
XX in physical mapping, in positional cloning of genes, in variety
XX identification and in evaluation of genetic variability within and
XX between plant tissues, populations, cultivars, species and species
XX groups. They may also be used to design hybridization probes for
XX oligonucleotide fingerprinting and library screening and to design
XX primers for microsatellite-primed PCR. Microsatellite markers are
XX useful to locate specific economically useful genes in plant genomes.
XX
XX Sequence 405 BP; 46 A; 175 C; 95 G; 89 T; 0 other;
XX
XX
Query Match 78.5%; Score 21.2; DB 21; Length 405;
Best Local Similarity 88.5%; Pred. No. 88;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 cccttgctgctctctctccct 27
DB 99 cccttgctgctctctctctct 124

```

```

KW Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 1; SEQ ID NO 1738; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 465 BP; 57 A; 167 C; 43 G; 198 T; 0 other;
XX
XX
Query Match 78.5%; Score 21.2; DB 22; Length 465;
Best Local Similarity 88.5%; Pred. No. 88;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 cccttgctgctctctctccct 27
DB 338 cccttgctgctctctctctct 363

```

```

RESULT 13
ABA43043
ID ABA43043 standard; DNA: 465 BP.
XX
XX
AC ABA43043;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human breast cell single exon nucleic acid probe #1738.
XX
XX

```

```

RESULT 14
ABA53458
ID ABA53458 standard; DNA: 465 BP.
XX
XX
AC ABA53458;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #1763.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX

```

```

XX Homo sapiens.
OS
XX
XX WO200157277-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 1763; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human fetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 465 BP; 57 A; 167 C; 43 G; 198 T; 0 other;
XX
SQ
Query Match 78.5%; Score 21.2; DB 22; Length 465;
Best Local Similarity 88.5%; Pred. No. 88;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 cccttgctgctctctctccct 27
Db 338 cccttctctctctctctctct 363

RESULT 15
ABA23228
ID ABA23228 standard; DNA; 465 BP.
XX
XX ABA23228;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #1694 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.

```

```

PR 26-MAY-2000; 2000US-0207456.
PR
PR 30-JUN-2000; 2000US-0608408.
PR
PR 03-AUG-2000; 2000US-0632366.
PR
PR 21-SEP-2000; 2000US-0234687.
PR
PR 27-SEP-2000; 2000US-0236359.
PR
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1; SEQ ID NO 1694; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 465 BP; 57 A; 167 C; 43 G; 198 T; 0 other;
XX
SQ
Query Match 78.5%; Score 21.2; DB 22; Length 465;
Best Local Similarity 88.5%; Pred. No. 88;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 cccttgctgctctctctccct 27
Db 338 cccttctctctctctctctct 363

Search completed: August 14, 2002, 22:06:53
Job time: 11708 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:57:31 ; Search time 203.42 Seconds
(Without alignments)
32.603 Million cell updates/sec

Title: US-09-707-919-9

Sequence: 1 cccctgcgtcctcctcctcctcctcct 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfilst1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.4	94.1	4481	4 US-09-041-886-18	Sequence 18, Appl
2	25	92.6	335	4 US-09-043-303-1	Sequence 1, Appl
3	25	92.6	623	4 US-09-043-303-5	Sequence 5, Appl
4	20.8	77.0	572	4 US-08-998-416-716	Sequence 716, Appl
5	19.2	71.1	189	2 US-08-733-505A-51	Sequence 51, Appl
6	19.2	71.1	189	2 US-08-733-505A-52	Sequence 52, Appl
7	19.2	71.1	189	2 US-08-733-505A-53	Sequence 53, Appl
8	19.2	71.1	189	2 US-08-733-505A-54	Sequence 54, Appl
9	19.2	71.1	944	1 US-08-665-617-1	Sequence 1, Appl
10	19.2	71.1	946	2 US-08-717-123-1	Sequence 1, Appl
11	19.2	71.1	1105	3 US-08-985-335-2	Sequence 2, Appl
12	19.2	71.1	1105	4 US-09-410-372-2	Sequence 1, Appl
13	19.2	71.1	2353	5 PCT-US92-06840-1	Sequence 1, Appl
14	19	70.4	48	4 US-08-979-608A-36	Sequence 36, Appl
15	19	70.4	84	4 US-08-979-608A-37	Sequence 37, Appl
16	19	70.4	300	4 US-09-135-994-3	Sequence 3, Appl
17	19	70.4	590	4 US-08-314-309A-10	Sequence 10, Appl
18	19	70.4	744	4 US-09-163-285-3	Sequence 3, Appl
19	19	70.4	1167	2 US-08-492-027A-5	Sequence 5, Appl
20	19	70.4	1362	4 US-08-979-608A-12	Sequence 12, Appl
21	19	70.4	1422	4 US-08-979-608A-13	Sequence 13, Appl
22	19	70.4	1512	4 US-09-163-285-1	Sequence 1, Appl
23	19	70.4	1617	4 US-08-979-608A-11	Sequence 11, Appl
24	19	70.4	1678	3 US-08-650-766-2	Sequence 2, Appl
25	19	70.4	1954	3 US-08-922-635-2	Sequence 12, Appl
26	19	70.4	2115	3 US-09-032-365A-12	Sequence 1, Appl
27	19	70.4	2116	1 US-08-701-380-1	Sequence 1, Appl

28	19	70.4	2150	2 US-08-861-464-13	Sequence 13, Appl
29	19	70.4	2150	4 US-08-396-001-13	Sequence 13, Appl
30	19	70.4	2150	4 US-09-323-433A-13	Sequence 4, Appl
31	19	70.4	2340	3 US-09-022-983-4	Sequence 3, Appl
32	19	70.4	2477	4 US-09-490-692-3	Sequence 1, Appl
33	19	70.4	3172	1 US-08-314-309A-1	Sequence 3, Appl
34	19	70.4	3318	3 US-08-650-766-3	Sequence 3, Appl
35	19	70.4	3318	3 US-08-922-635-3	Sequence 1, Appl
36	19	70.4	3385	3 US-08-922-635-1	Sequence 1, Appl
37	19	70.4	3385	3 US-08-922-635-1	Sequence 1, Appl
38	19	70.4	5057	3 US-08-651-999A-6	Sequence 6, Appl
39	19	70.4	5057	4 US-09-385-752-6	Sequence 21, Appl
40	19	70.4	15202	3 US-08-922-635-21	Sequence 2, Appl
41	19	70.4	15378	3 US-08-785-420-1	Sequence 2, Appl
42	19	70.4	53526	3 US-08-658-136-2	Sequence 1, Appl
43	19	70.4	53577	2 US-08-658-136-1	Sequence 7, Appl
44	18.6	68.9	3527	2 US-08-909-965C-7	Sequence 3, Appl
45	18.6	68.9	11703	4 US-09-101-886B-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-18
; Sequence 18, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..4099
; US-09-041-886-18

Query Match 94.1% Score 25.4; DB 4; Length 4481;
Best Local Similarity 96.3% Pred. No. 0.36;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcctctcccccct 27
|||||
Db 468 ccccttcgctgcctctcccccct 494

RESULT 2
US-09-043-303-1
; Sequence 1, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(355)
US-09-043-303-1

Query Match 92.6%; Score 25; DB 4; Length 355;
Best Local Similarity 92.6%; Pred. No. 0.46;
Matches 25; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcctctcccccct 27
|||||
Db 166 cccctcgctgcctctcccccct 192

RESULT 3
US-09-043-303-5
; Sequence 5, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (341)..(583)
; FEATURE:
; OTHER INFORMATION: TSP-2
US-09-043-303-5

Query Match 92.6%; Score 25; DB 4; Length 623;
Best Local Similarity 92.6%; Pred. No. 0.47;
Matches 25; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcctctcccccct 27

Db 166 cccctcgctgcctctcccccct 192
|||||

RESULT 4
US-08-998-416-716
; Sequence 716, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCCI976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 716:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 572 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1469UP
US-08-998-416-716

Query Match 77.0%; Score 20.8; DB 4; Length 572;
Best Local Similarity 91.7%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 cctcgctgcctctcccccct 27
|||||
Db 141 cctcgctgcctctcccccct 164

RESULT 5
US-08-733-505A-51/C
; Sequence 51, Application US/08733505A
; Patent No. 5856445
; GENERAL INFORMATION:
; APPLICANT: KORSMEYER, STANLEY J.
; TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF

TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,505A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 965458
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-6092
TELEFAX: (314) 727-5188
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 189 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-733-505A-51

Query Match 71.1%; Score 19.2; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cccttgctgctcctcctccccc 25
|||||
DB 69 CCCTTGCTGCTCCTCGTCCCGC 46

RESULT 6
US-08-733-505A-52/C
Sequence 52, Application US/08733505A
Patent No. 5856445
GENERAL INFORMATION:
APPLICANT: KORSMEYER, STANLEY J.
TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF
TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,505A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 965458

TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 189 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-733-505A-52

Query Match 71.1%; Score 19.2; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cccttgctgctcctcctccccc 25
|||||
DB 69 CCCTTGCTGCTCCTCGTCCCGC 46

RESULT 7
US-08-733-505A-53/C
Sequence 53, Application US/08733505A
Patent No. 5856445
GENERAL INFORMATION:
APPLICANT: KORSMEYER, STANLEY J.
TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF
TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,505A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 965458
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 727-5188
TELEFAX: (314) 727-6092
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 189 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-733-505A-53

Query Match 71.1%; Score 19.2; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cccttgctgctcctcctccccc 25
|||||
DB 69 CCCTTGCTGCTCCTCGTCCCGC 46

RESULT 8
US-08-733-505A-54/C
; Sequence 54, Application US/08733505A
; Patent No. 5856445
; GENERAL INFORMATION:
; APPLICANT: KORMEYER, STANLEY J.
; TITLE OF INVENTION: SERINE SUBSTITUTED MUTANTS OF
; TITLE OF INVENTION: BCL-XL/BCL-2 ASSOCIATED CELL DEATH REGULATOR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733.505A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 965458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 189 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-733-505A-54

Query Match 71.1%; Score 19.2; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccttcgtcgtcctcctcctccccc 25
|||||
Db 69 CCCTTCGTCTCCTCCGTCCTCCGCGC 46

RESULT 9
US-08-665-617-1/C
; Sequence 1, Application US/08665617
; Patent No. 5663316
; GENERAL INFORMATION:
; APPLICANT: Xudong, Yin
; TITLE OF INVENTION: Gene and Protein for Regulation of Cell Death
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665.617

; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: CL-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-665-617-1

Query Match 71.1%; Score 19.2; DB 1; Length 944;
Best Local Similarity 87.5%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccttcgtcgtcctcctcctccccc 25
|||||
Db 315 CCCTTCGTCTCCTCCGTCCTCCGCGC 292

RESULT 10
US-08-717-123-1/C
; Sequence 1, Application US/08717123
; Patent No. 5965703
; GENERAL INFORMATION:
; APPLICANT: Horne, William A.
; TITLE OF INVENTION: Oltersdorf, Tilmann
; TITLE OF INVENTION: Human BAD Polypeptides, Encoding Nucleic
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/717.123
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-ID 1929
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 52..555
; US-08-717-123-1

Query Match 71.1%; Score 19.2; DB 2; Length 946;
Best Local Similarity 87.5%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccttcgtcgtcctcctccccc 25
|||||
DB 306 CCTTCGTCGTCCTCCGCCGCCG 283

RESULT 11

US-08-985-335-2/c
Sequence 2, Application US/08985335
Patent No. 6080847

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Yue, Henry

APPLICANT: Lal, Preeti

APPLICANT: Shah, Purni

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/985,335

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0421 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-845-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1105 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: 358673

CLONE: SYNORAB01

US-08-985-335-2

Query Match 71.1%; Score 19.2; DB 3; Length 1105;

Best Local Similarity 87.5%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccttcgtcgtcctcctccccc 25
|||||
DB 468 CCTTCGTCGTCCTCCGCCGCCG 445

RESULT 12

US-09-410-372-2/c

Sequence 2, Application US/09410372

Patent No. 6281334

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Yue, Henry

APPLICANT: Lal, Preeti

APPLICANT: Shah, Purni

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/410,372

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/985,335

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0421 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-845-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1105 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: 358673

CLONE: SYNORAB01

US-09-410-372-2

Query Match 71.1%; Score 19.2; DB 4; Length 1105;

Best Local Similarity 87.5%; Pred. No. 53;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccttcgtcgtcctcctccccc 25
|||||
DB 468 CCTTCGTCGTCCTCCGCCGCCG 445

RESULT 13

PCT-US92-06840-1/c

Sequence 1, Application PC/TUS9206840

GENERAL INFORMATION:

APPLICANT: Shi, Yang

APPLICANT: Seto, Edward

APPLICANT: Shenk, Thomas

TITLE OF INVENTION: YY1 TRANSCRIPTION FACTOR AND METHODS OF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerd & Soffen

STREET: 1180 Avenue of the Americas - 7th Floor

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

Thu Aug 15 09:03:20 2002

us-09-707-919-9.rni

Page 7

1 FILING DATE: 27-NOV-1996
 2 ATTORNEY/AGENT INFORMATION:
 3 NAME: Myers, Louis
 4 REGISTRATION NUMBER: 35,965
 5 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 398J/59818)
 6 TELECOMMUNICATION INFORMATION:
 7 TELEPHONE: 617/542-5070
 8 TELEFAX: 617/542-8906
 9 INFORMATION FOR SEQ ID NO: 37:
 10 SEQUENCE CHARACTERISTICS:
 11 LENGTH: 84 base pairs
 12 TYPE: nucleic acid
 13 STRANDEDNESS: single
 14 TOPOLOGY: linear
 15 FEATURE:
 16 NAME/KEY: Coding Sequence
 17 LOCATION: 1...84
 18 SEQUENCE DESCRIPTION: SEQ ID NO: 37:
 19 GS-08-979-608A-37

```

Query Match      70.4%: Score 19: DB 4: Length 84:
Best Local Similarity 81.5%: Pred. No. 57:
Matches 22: Conservative 0: Mismatches 5: Indels 0: Gaps 0:

OY 1 ccccttcgctgctctctctctccccc 27
    ||| ||||| ||||| ||||| |||||
db 34 ccttcgctgctctctctctctctctcc 8

```

Search completed: August 14, 2002, 21:57:32
Job time: 13885 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:36 ; Search time 7749.14 Seconds
(without alignments)

47.027 Million cell updates/sec

Title: US-09-707-919-9

Perfect score: 27

Sequence: 1 cccctcgtcctcctcctcctccccc 27

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	27	100.0	1100	10	BM455214	BM455214 AGENCOURT
2	26	96.3	126	10	F14808	F14808 SSC20D02 PO
3	25.4	94.1	482	9	AL039573	AL039573 DFE2P434D
4	25.4	94.1	500	10	BI547486	BI547486 603191091
5	22.8	84.4	998	12	CNS04RSZ	AL304436 Tetradon
6	22.2	82.2	397	10	BU185471	BU185471 BUI85471
7	22.2	82.2	555	10	BF618652	BF618652 HVSMEC000
8	22.2	82.2	670	9	AL135709	AL135709 DFE2P434A
9	22.2	82.2	709	12	BH058331	BH058331 RPCI-24-3
10	22.2	82.2	809	9	AV710405	AV710405 AV710405
11	22.2	82.2	1096	10	BG826523	BG826523 602750392
12	22.2	82.2	1132	12	AG073208	AG073208 Pan trogl
13	22.2	82.2	1201	12	CNS0164L	AL106287 Drosoph11
14	21.8	80.7	477	9	BB822913	BB822913 BB822913
15	21.8	80.7	724	12	AG089170	AG089170 Pan trogl
16	21.4	79.3	405	9	AW397925	AW397925 sg69h11.y
17	21.4	79.3	658	12	AG688040	AG688040 ndxd0076j

18	21.2	78.5	211	9	AL603433	AL603433 DFE2P686H
19	21.2	78.5	248	9	AA732029	AA732029 n287a02.s
20	21.2	78.5	284	10	W35372	W35372 zc07h01.s1
21	21.2	78.5	297	10	BF223384	BF223384 7687f11.x
22	21.2	78.5	306	9	AA470994	AA470994 x280a04.x
23	21.2	78.5	309	10	BE550941	BE550941 7d66c06.x
24	21.2	78.5	317	9	AA770579	AA770579 h166h10.x
25	21.2	78.5	323	9	AA766840	AA766840 oc87e05.s
26	21.2	78.5	328	10	BE670764	BE670764 7604f05.x
27	21.2	78.5	329	10	BE674474	BE674474 7604f05.x
28	21.2	78.5	364	9	AT500389	AT500389 tm96a11.x
29	21.2	78.5	371	9	AA824422	AA824422 oc78a10.s
30	21.2	78.5	376	9	AA533044	AA533044 n160b07.s
31	21.2	78.5	379	10	R79172	R79172 y184d04.s1
32	21.2	78.5	385	12	AO911478	AO911478 LMAFEV1.1
33	21.2	78.5	403	9	AA207344	AA207344 UI-H-B11-
34	21.2	78.5	414	9	AA969801	AA969801 o082e10.s
35	21.2	78.5	417	12	A2724018	A2724018 RPCI-24-6
36	21.2	78.5	421	12	AO881832	AO881832 HS_5273_A
37	21.2	78.5	424	9	AA293888	AA293888 zt61b12.r
38	21.2	78.5	437	9	AA743640	AA743640 ny24901.s
39	21.2	78.5	437	9	AT352271	AT352271 qz12c01.x
40	21.2	78.5	449	10	BG550798	BG550798 sad91g01.
41	21.2	78.5	454	9	AT459045	AT459045 t396d03.x
42	21.2	78.5	458	9	AA043094	AA043094 zK35a09.s
43	21.2	78.5	460	10	R79927	R79927 y191a02.s1
44	21.2	78.5	461	9	AA807682	AA807682 nv66908.s
45	21.2	78.5	469	10	D82418	D82418 HUMBCA4566

ALIGNMENTS

RESULT 1
LOCUS BM455214 1100 bp mRNA linear EST 05-PEB-2002
DEFINITION AGENCOURT 6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
5', mRNA sequence.
ACCESSION BM455214
VERSION BM455214.1 GI:18504254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/.
1 (bases 1 to 1100)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12134 row: k column: 12
High quality sequence stop: 623.

FEATURES

source
1..1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

size-selected for average insert size 2.5 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NIHRI, National Institutes of Health). Note: this is a NIH-MGC Library.*

BASE COUNT

57 a 222 c 150 g 71 t

ORIGIN

Query Match 94.1%; Score 25.4; DB 10; Length 500;
Best Local Similarity 96.3%; Pred. No. 3.5e+02;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ccccttcgtcgtccttcctccct 27
|||||

Db 118 ccccttcgtcgtccttcctccct 144

RESULT 5

CNS04RSZ/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

Db 713 ccccttcgtcgtccttcctccct 688

RESULT 6

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

24; Conservative

0; Mismatches

3; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

0; Gaps

0; Indels

Bj185471 397 bp mRNA linear EST 24-JUN-2002

Bj185471 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens

cDNA clone pphb7m03 5', mRNA sequence.

Bj185471 1 GI:18353416

EST.

Physcomitrella patens subsp. patens.

Physcomitrella patens subsp. patens

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Bryophyta;

Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

1 (bases 1 to 397)

Fujita, T., Shinozaki, K., Seki, M., Kamaya, A., Uchiyama, I., Nishiyama, T.,

Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe

, M.

Comparison of the moss Physcomitrella patens genome with flowering

plants genome

Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo

excised from a modified lps phase vector (Mo bi Tec, Germany). XhoI

digested 5' end of cDNA is ligated to SalI site of the vector, and

the BamHI digested 3' end including poly-A tail is ligated to BamHI

site of the vector. cDNA insert could be amplified with

conventional T7 and T3 primers. This normalized full-length cDNA

library was generated basically according to the method described

in Genome Research 10, 1617-1630 (2000). Carninci, P. et al.

Protonemata were blended by the POLYTRON, and then cultivated on

the BCD medium containing 0.5mM BA (benzylaminopurine) for 8 to 13

days under the continuous light.

Location/Qualifiers

1..397

/organism="Physcomitrella patens subsp. patens"

/db_xref="taxon:145481"

/clone="pphb7m03"

/clone_lib="normalized full length cDNA library,

chloronemata, caulonemata and malformed buds"

/tissue-type="mixture of chloronemata, caulonemata and

malformed buds"

malformed buds"

malformed buds"

ORGANISM	Hordeum vulgare			
REFERENCE	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae 1 (bases 1 to 555)			
AUTHORS	Wing,R., Close,T.J., Kleinbofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D.			
TITLE	Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling shoot cDNA library unpublished (2001)			
JOURNAL	On Dec 18, 2000 this sequence version replaced gi:11882386.			
COMMENT	Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: twing@clemson.edu Total hg bases = 242 Seq primer: AATTAACTCTCAGCTAAGGC High quality sequence stop: 455.			
FEATURES	Location/Qualifiers			
Source	1..555			
	/organism="Hordeum vulgare"			
	/cultivar="Morex"			
	/db_xref="taxon:4513"			
	/clone="HVSMEC0007B15f"			
	/clone_lib="Hordeum vulgare seedling shoot EST library			
	HVCDDM0003 (Etiolated and unstressed)"			
	/issue_type="Seedling shoot"			
	/lab_host="TUC121"			
	/note="Vector: lambdaZAP, Site_1: EcoRI, Site_2: XhoI; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and ceftaxime in covered crystallization dishes. Five-day old seedling shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million plu were in vivo excised to give plusescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the university of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see			
	http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinbofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"			
BASE COUNT	126 a 137 c 226 g 66 t			
ORIGIN				
Query Match	82.2%; Score 22.2; DB 10; Length 555;			
Best Local Similarity	88.9%; Pred. No. 2.9e+03;			
Matches	24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
0y	1 cccctcgatgcctcctcctccct 27			
Db	547 cccctcgatgcctcctcctccct 521			
RESULT	8			
LOCUS	AL135709 670 bp mRNA linear EST 25-FEB-2000			

DEFINITION	DKFZp434A152.r1.434 (synonym: htes3) Homo sapiens cDNA clone			
ACCESSION	DKFZp434A152.5', mRNA sequence.			
VERSION	AL135709			
KEYWORDS	AL135709.1 GI:6603896			
SOURCE	EST.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	1 (bases 1 to 670)			
COMMENT	Ansoerge,W., Wirtner,U., Mewes,H.W., Gassenhuber,J. and Wiemann,S. EST (Ansoerge, et al.) Unpublished (1999) Contact: Ansoerge W MIPS			
FEATURES	Am Klopferspitz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert This clone (DKFZp434A152) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Location/Qualifiers 1..670 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZp434A152" /clone_1lb="434 (synonym: htes3)" /tissue_type="testis" /dev_stage="adult" /lab_host="DH10B" /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"			
BASE COUNT	175 a 170 c 216 g 109 t .			
ORIGIN				
Query Match	82.2%; Score 22.2; DB 9; Length 670;			
Beat Local Similarity	88.9%; Pred. No. 2.9e+03;			
Matches	24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1 cccctgcgtcgtcctcctcctccct 27 			
Db	92 CCCCTTCCTTCGTCCTCTCCTCCT 66			
RESULT	9			
BH058331/c	BH058331 709 bp DNA linear GSS 18-JUL-2001			
LOCUS	RPCT-24-326F19.TV RPCT-24 Mus musculus genomic clone RPCT-24-326F19			
DEFINITION	, DNA sequence.			
ACCESSION	BH058331			
VERSION	BH058331.1 GI:14867206			
KEYWORDS	GSS.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 709)			
JOURNAL	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E., Russell,D., de Jong,P. and Fraser,C.M.			
COMMENT	Mouse BAC End Sequences from Library RPCT-24 Unpublished (1999) Other_GSSs: RPCT-24-326F19.TJ Contact: Shaying Zhao Department of Eukaryotic Genomics 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200			

ACCESSION AG073208
 VERSION GI:16625010
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC Library clone:PTB-064024.R.
 ORGANISM Pan troglodytes

REFERENCE 1 (sites)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1132)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbes@gs.c.riken.go.jp, URL: http://hqp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.
 PRIMERS

Sequencing: M13Rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI.
 R.Site 2 : SacI.
 Location/Qualifiers
 1..1132
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-064024.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC library"
 BASE COUNT 363 a 302 c 442 g 17 t 8 others
 ORIGIN

Query Match 82.2%; Score 22.2; DB 12; Length 1132;
 Best Local Similarity 88.9%; Pred. No. 3.1e+03;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ccccttgctgcctctctcccccct 27
 ||||| ||||| ||||| ||||| |||||
 Db 259 CCTCTTCCTCCTCCTCTCTCTCCCT 233

RESULT 13
 CNS0164L 1201 bp DNA linear GS. 26-JUL-1999
 LOCUS CNS0164L/C
 DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN15CJ1 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL106287
 AL106287.1 GI:5621177
 VERSION AL106287.1
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1201)
 GENOSCOPE.
 DIRECT SUBMISSION
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segr@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebl.ac.uk. This Drosophila melanogaster BAC library (Dros BAC) was made up by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelBAC11.

FEATURES
 source
 1..1201
 /organism="Drosophila melanogaster"
 /plasmid="pBelBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN15CJ1"
 /note="end : SP6"
 BASE COUNT 304 a 204 c 230 g 275 t 188 others
 ORIGIN

Query Match 82.2%; Score 22.2; DB 12; Length 1201;
 Best Local Similarity 77.8%; Pred. No. 3.1e+03;
 Matches 21; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccccttgctgcctctctcccccct 27
 ||||| ||||| ||||| ||||| |||||
 Db 1032 CCCCTTCTCTCTCTCTCTCTCCCT 1006

RESULT 14
 BB822913 477 bp mRNA linear EST 19-NOV-2001
 LOCUS BB822913
 DEFINITION BB822913 RIKEN full-length enriched, mammary gland RCB-0526 Jyg-MC(A) cDNA Mus musculus cDNA clone G830025B02.3', mRNA sequence.
 ACCESSION BB822913
 VERSION BB822913
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 REFERENCE 1 (bases 1 to 477)
 AUTHORS Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiroaka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishikawa, K., Nomura, K., Nunasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sekazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watanabe, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Yoshinori Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.c.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res.

10 (11), 1757-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source location/Qualifiers
 1.477
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="G830025B02"
 /clone.lib="RIKEN full-length enriched, mammary gland
 RCB-0526 Jy9-MC(A) cDNA"
 /tissue_type="mammary gland"
 /cell_line="RCB-0526 Jy9-MC(A)"
 BASE COUNT 101 a 134 c 124 g 118 t
 ORIGIN

Query Match 80.7%; Score 21.8; DB 9; Length 477;
 Best Local Similarity 92.0%; Pred. No. 3.7e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 cccttcgctgcctcctcctccccc 26
 |||||||
 Db 186 CCCTTCGTCCTCCTCCTCCACC 210

RESULT 15

AG089170 724 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-088F13.R, genomic survey sequence.
 DEFINITION AG089170
 ACCESSION AG089170.1 GI:16640972
 VERSION
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male
 BAC Library clone:PTB-088F13.R.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of Library PTB
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 724)
 REFERENCE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 JOURNAL Direct Submission
 TITLE Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@gs.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS
 Sequencing: M13Rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..724
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-088F13.R"
 /sex="male"
 /cell_type="lymphoblast"

FEATURES

source
 1..724
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-088F13.R"
 /sex="male"
 /cell_type="lymphoblast"

BASE COUNT 136 a 258 c 193 g 132 t 5 others
 ORIGIN

Query Match 80.7%; Score 21.8; DB 12; Length 724;
 Best Local Similarity 92.0%; Pred. No. 3.9e+03;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 cccttcgctgcctcctcctccccc 25
 |||||||
 Db 6 CCCCTTCGTCCTCCTCCTCCACC 30

Search completed: August 14, 2002, 21:04:43
 Job time: 11031 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:49:05 ; Search time 2563.92 Seconds
(without alignments)
220.372 Million cell updates/sec

Title: US-09-707-919-10

Perfect score: 27
Sequence: 1 cccctcgtcgtcgtccttcctccct 27

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	27	100.0	4163	9	HSDANSCA2	Y08262 H. sapiens m
2	27	100.0	4200	6	AE2706	AE2706 Sequence 7
3	27	100.0	4481	6	ARI53580	ARI53580 Sequence
4	27	100.0	4481	6	HSU70323	070323 Human ataxi
5	26.6	98.5	355	6	ARI59544	ARI59544 Sequence
6	26.6	98.5	572	6	ARI59558	ARI59558 Sequence
7	26.6	98.5	623	6	ARI59546	ARI59546 Sequence
8	25.4	94.1	264	9	AF330032	AF330032 Papio ham
9	25.4	94.1	384	9	AF330030	AF330030 Presbytis
10	25.4	94.1	390	9	AF330028	AF330028 Pan trogl
11	25.4	94.1	409	9	AF330029	AF330029 Gorilla g
12	25.4	94.1	409	2	AC004085	AC004085 Homo sapi
13	24.4	90.4	303	9	AF330031	AF330031 Macaca mu
14	24.4	90.4	322	9	AF330033	AF330033 Macaca ra
15	22.8	84.4	104481	2	AP003844	AP003844 Oryza sat
16	22.8	84.4	151311	2	AP004262	AP004262 Oryza sat
17	22.2	82.2	699	9	HSB329274	AJ329274 Homo sapi
18	22.2	82.2	2487	9	AK057572	AK057572 Homo sapi
19	22.2	82.2	153116	8	AP003392	AP003392 Oryza sat
20	22.2	82.2	169540	8	AC009248	AC009248 Homo sapi
21	22.2	82.2	180508	2	AC096194	AC096194 Rattus no
22	21.8	80.7	166070	2	AC090871	AC090871 Oryza sat
23	21.8	80.7	229896	14	AF232689	AF232689 Rat cytom
24	21.2	78.5	41944	3	AC005929	AC005929 Leishmani
25	21.2	78.5	120116	8	AC092390	AC092390 Oryza sat
26	21.2	78.5	146240	2	AC103012	AC103012 Rattus no
27	21.2	78.5	160922	2	AC104848	AC104848 Oryza sat
28	21.2	78.5	176186	2	AP003335	AP003335 Oryza sat
29	20.8	77.0	610	8	AB047923	AB047923 Oryza sat
30	20.8	77.0	62070	8	NC2E4	AL45102 Neurospor
31	20.8	77.0	148608	8	H0711G06	AL442115 Oryza sat
32	20.8	77.0	170659	2	OSJN00025	AL606527 Oryza sat
33	20.8	77.0	171075	2	OSJN00052	U14077 H. sapiens m
34	20.6	76.3	1559	5	HSY1NPE1	U37272 Gallus gall
35	20.6	76.3	1708	5	GSU37272	M74590 Mouse delta
36	20.6	76.3	1988	10	MUSDELTA	M76541 Human DNA-b
37	20.6	76.3	2176	9	HDNFEIYD	AK056159 Homo sapi
38	20.6	76.3	2233	9	AK056159	M73963 Mus musculu
39	20.6	76.3	2330	10	MUSUCRBP	M77698 Homo sapien
40	20.6	76.3	2353	9	HUMCRP	AF326769 Mus muscu
41	20.6	76.3	2939	10	AF326769	L13969 Mouse delta
42	20.6	76.3	3041	10	MOSTRANS01	AF023910 Physarum
43	20.6	76.3	3325	3	AF023910	AF435838 Drosophila
44	20.6	76.3	3502	3	AF435838	AK056477 Homo sapi
45	20.6	76.3	3908	9	AK056477	

ALIGNMENTS

RESULT 1	HSDANSCA2	4163 bp	mRNA	linear	PRI 09-JAN-1997
LOCUS	HSDANSCA2				
DEFINITION	H. sapiens mRNA for SCA2 protein.				
ACCESSION	Y08262				
VERSION	Y08262.1 GI:1770389				
KEYWORDS	SCA2 gene.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
	1 (bases 1 to 4163)				
	Imbert,G., Saudou,F., Yvert,G., Devys,D., Trotter,Y.,				
	Garnier,J.M., Weber,C., Mandel,J.L., Cancel,G., Abbas,N., Duerr,A.,				
	Didierjean,O., Stevanin,G., Agid,Y. and Brice,A.				
	Cloning of the gene for spinocerebellar ataxia 2 reveals a locus				
	with high sensitivity to expanded CAG/glutamine repeats				
	Nat. Genet. 14 (3), 285-291 (1996)				
TITLE	JOURNAL				
MEDLINE	2 (bases 1 to 4163)				
REFERENCE	97051922				
AUTHORS	Imbert,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-SEP-1996) G. Imbert, I.G.B.M.C., Departement Of				


```
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecoidea; Papio.
REFERENCE
  1 (bases 1 to 264)
AUTHORS
  Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
  Brahmachari,S.K.
TITLE
  CAG repeat instability at SCA2 locus: anchoring CAA interruptions
  and linked single nucleotide polymorphisms
JOURNAL
  Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
REFERENCE
  2 (bases 1 to 264)
AUTHORS
  Choudhry,S. and Brahmachari,S.K.
TITLE
  Direct Submission
JOURNAL
  Submitted (21-DEC-2000) Functional Genomics Unit, Center for
  Biochemical Technology, Delhi University Campus, Mall Road, Delhi
  110 007, India
FEATURES
  source
    1..264
    /organism="Papio hamadryas"
    /db_xref="taxon:9557"
    <1..>264
    /gene="SCA2"
    /note="spino cerebellar ataxia 2"
BASE COUNT
  25 a 130 c 78 g 31 t
ORIGIN
Query Match
  94.1%; Score 25.4; DB 9; Length 264;
Best Local Similarity 96.3%; Pred. No. 17;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ccccttcgctgcgtccttcctccct 27
Db 20 CCCCTTCGTCGTCCTTCCTCCCT 46
RESULT 9
AF330030 384 bp DNA linear PRI 08-NOV-2001
LOCUS
  Presbylis entellus SCA2 gene, partial sequence.
DEFINITION
  AF330030
ACCESSION
  AF330030.1 GI:12382832
KEYWORDS
  Hanuman langur.
  Presbylis entellus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
  Colobinae; Presbylis.
REFERENCE
  1 (bases 1 to 384)
AUTHORS
  Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
  Brahmachari,S.K.
TITLE
  CAG repeat instability at SCA2 locus: anchoring CAA interruptions
  and linked single nucleotide polymorphisms
JOURNAL
  Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
REFERENCE
  2 (bases 1 to 384)
AUTHORS
  Choudhry,S. and Brahmachari,S.K.
TITLE
  Direct Submission
JOURNAL
  Submitted (21-DEC-2000) Functional Genomics Unit, Center for
  Biochemical Technology, Delhi University Campus, Mall Road, Delhi
  110 007, India
FEATURES
  source
    1..384
    /organism="Presbylis entellus"
    /db_xref="taxon:9574"
    <1..>384
    /gene="SCA2"
    /note="spino cerebellar ataxia 2"
BASE COUNT
  46 a 178 c 109 g 51 t
ORIGIN
Query Match
  94.1%; Score 25.4; DB 9; Length 384;
Best Local Similarity 96.3%; Pred. No. 16;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ccccttcgctgcgtccttcctccct 27
Db 14 CCCCTTCGTCGTCCTTCCTCCCT 40
RESULT 10
AF330028 390 bp DNA linear PRI 08-NOV-2001
LOCUS
  Pan troglodytes SCA2 gene, partial sequence.
DEFINITION
  AF330028
ACCESSION
  AF330028.1 GI:12382830
KEYWORDS
  chimpanzee.
  Pan troglodytes
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE
  1 (bases 1 to 390)
AUTHORS
  Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
  Brahmachari,S.K.
TITLE
  CAG repeat instability at SCA2 locus: anchoring CAA interruptions
  and linked single nucleotide polymorphisms
JOURNAL
  Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
REFERENCE
  2 (bases 1 to 390)
AUTHORS
  Choudhry,S. and Brahmachari,S.K.
TITLE
  Direct Submission
JOURNAL
  Submitted (21-DEC-2000) Functional Genomics Unit, Center for
  Biochemical Technology, Delhi University Campus, Mall Road, Delhi
  110 007, India
FEATURES
  source
    1..390
    /organism="Pan troglodytes"
    /db_xref="taxon:9598"
    <1..>390
    /note="microsatellite"
    /rpt_type="tandem"
    /rpt_unit="cag"
    <1..>390
    /gene="SCA2"
    /note="spino cerebellar ataxia 2"
BASE COUNT
  48 a 183 c 110 g 49 t
ORIGIN
Query Match
  94.1%; Score 25.4; DB 9; Length 390;
Best Local Similarity 96.3%; Pred. No. 16;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ccccttcgctgcgtccttcctccct 27
Db 14 CCCCTTCGTCGTCCTTCCTCCCT 40
RESULT 11
AF330029 409 bp DNA linear PRI 08-NOV-2001
LOCUS
  Gorilla gorilla SCA2 gene, partial sequence.
DEFINITION
  AF330029
ACCESSION
  AF330029.1 GI:12382831
KEYWORDS
  gorilla.
  Gorilla gorilla
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
REFERENCE
  1 (bases 1 to 409)
AUTHORS
  Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
  Brahmachari,S.K.
TITLE
  CAG repeat instability at SCA2 locus: anchoring CAA interruptions
  and linked single nucleotide polymorphisms
JOURNAL
  Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
11689490
```

REFERENCE 2 (bases 1 to 409)
 AUTHORS Choudhry, S. and Brahmachari, S. K.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-2000) Functional Genomics Unit, Center for
 Biochemical Technology, Delhi University Campus, Mall Road, Delhi
 110 007, India
 FEATURES Location/Qualifiers
 source 1..409
 /organism="Gorilla gorilla"
 /db_xref="taxon:9593"
 <!.>.409
 /gene="SCA2"
 /note="spinocerebellar ataxia 2"
 BASE COUNT 35 a 196 c 120 g 58 t
 ORIGIN
 Query Match 94.1%; Score 25.4; DB 9; Length 409;
 Best Local Similarity 96.3%; Pred. No. 16;
 Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 cccctcgcgtcgtcctccctcct 27
 Db 48 CCCCTTCGTGCTCCTCTCTCCCT 74
 RESULT 12
 AC004085/C
 LOCUS AC004085.6
 DEFINITION Homo sapiens clone RP11-42B1, WORKING DRAFT SEQUENCE, 20 unordered
 pieces.
 AC004085 GI:11079383
 VERSION AC004085.6
 KEYWORDS HTG: PHASE1; HTGS: DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 231758)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alshrocks, S.L., Anaratunga, H.C., Are, J.R., Banks, T., Barbata, J.,
 Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulvik, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Lewis, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W.,
 Louieged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
 Meitzner, M., Miner, C., Miner, Z., Mitchell, T., Mohabadi, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S.,
 Ogulu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
 Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshari, N.,
 Stinson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, F., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellrod, B., Thomas, N.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Wellington, S., Williams, G., Williamson, A., Wleczyk, R., Wood, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.
 and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 231758)
 Morley, K.C.
 Submitted (30-JAN-1998) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 3, 2000 this sequence version replaced gi:1966929.
 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc.help@bcm.tmc.edu
 Project Information
 Center project name: UG
 Center clone name: RP11-42B1

----- Summary Statistics -----
 Assembly program: Phrap; version 0.990329
 Consensus quality: 224788 bases at least Q40
 Consensus quality: 229074 bases at least Q30
 Consensus quality: 230948 bases at least Q20
 Estimated insert size: 227237; sum-of-contigs estimation
 Estimated insert size: 317311; agarose-gel estimation
 Quality coverage: 6.3x in Q20 bases; agarose-gel estimation
 Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 33241: contig of 33241 bp in length
 33242 33241: gap of unknown length
 33342 33341: contig of 23050 bp in length
 56391 56391: gap of unknown length
 56491 56491: gap of unknown length
 56492 81323: contig of 24832 bp in length
 81324 81423: gap of unknown length
 81424 102538: contig of 2115 bp in length
 102539 102638: gap of unknown length
 102639 119710: contig of 17072 bp in length
 119711 119810: gap of unknown length
 119811 136913: contig of 17103 bp in length
 136914 137013: gap of unknown length
 153285 153285: contig of 16272 bp in length
 153286 153385: gap of unknown length
 153386 167987: contig of 14602 bp in length
 167988 168087: gap of unknown length
 168088 178731: contig of 10644 bp in length
 178732 178831: gap of unknown length
 178832 186641: contig of 7810 bp in length
 186642 186741: gap of unknown length
 186742 193215: contig of 6474 bp in length
 193216 193315: gap of unknown length
 193316 201310: contig of 7995 bp in length
 201311 201410: gap of unknown length
 201411 208647: contig of 7237 bp in length
 208648 208747: gap of unknown length
 213802 213802: contig of 5055 bp in length
 213803 213902: gap of unknown length
 218049 218049: contig of 4147 bp in length
 218050 218149: gap of unknown length
 223316 223316: contig of 5167 bp in length
 223317 223416: gap of unknown length

```

FEATURES
SOURCE
1..231758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-42B1"

BASE COUNT      64974 a 51086 c 51148 g 62641 t 1909 others
ORIGIN

Query Match      94.1%; Score 25.4; DB 2; Length 231758;
Best Local Similarity 96.3%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
Db 89318 CCCCTTCGTCGTCCTTCCTCCCT 89292

RESULT 13
AF330031      303 bp  DNA      linear  PRI 08-NOV-2001
LOCUS      Macaca mulatta SCA2 gene, partial sequence.
ACCESSION  AF330031
VERSION    AF330031.1 GI:12382833
KEYWORDS

SOURCE
.
Rhesus monkey.
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 303)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
JOURNAL
11689490
2 (bases 1 to 303)
Choudhry,S. and Brahmachari,S.K.
Direct Submission
Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
SOURCE
1..303
Location/Qualifiers
/organism="Macaca mulatta"
/db_xref="taxon:9544"
<1..>303
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT      32 a 143 c 92 g 36 t
ORIGIN

Query Match      90.4%; Score 24.4; DB 9; Length 303;
Best Local Similarity 96.2%; Pred. No. 39;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccccc 26
|||||
Db 14 CCCCTTCGTCGTCCTTCCTCCCC 39

RESULT 14
AF330033      322 bp  DNA      linear  PRI 08-NOV-2001
LOCUS      Macaca mulatta SCA2 gene, partial sequence.
ACCESSION  AF330033
VERSION    AF330033.1 GI:12382833
KEYWORDS
SOURCE
.
Rhesus monkey.
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 322)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
JOURNAL
11689490
2 (bases 1 to 322)
Choudhry,S. and Brahmachari,S.K.
Direct Submission
Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
SOURCE
1..322
Location/Qualifiers
/organism="Macaca radiata"
/db_xref="taxon:9548"
<1..>322
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT      32 a 155 c 95 g 40 t
ORIGIN

```

```

DEFINITION  Macaca radiata SCA2 gene, partial sequence.
ACCESSION  AF330033
VERSION    AF330033.1 GI:12382833
KEYWORDS
SOURCE
.
bonnet macaque.
Macaca radiata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 322)
Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and
Brahmachari,S.K.
CAG repeat instability at SCA2 locus: anchoring CAA interruptions
and linked single nucleotide polymorphisms
Hum. Mol. Genet. 10 (21), 2437-2446 (2001)
JOURNAL
11689490
2 (bases 1 to 322)
Choudhry,S. and Brahmachari,S.K.
Direct Submission
Submitted (21-DEC-2000) Functional Genomics Unit, Center for
Biochemical Technology, Delhi University Campus, Mall Road, Delhi
110 007, India
FEATURES
SOURCE
1..322
Location/Qualifiers
/organism="Macaca radiata"
/db_xref="taxon:9548"
<1..>322
/gene="SCA2"
/note="Spinocerebellar ataxia 2"

BASE COUNT      32 a 155 c 95 g 40 t
ORIGIN

Query Match      90.4%; Score 24.4; DB 9; Length 322;
Best Local Similarity 96.2%; Pred. No. 38;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccccc 26
|||||
Db 43 CCCCTTCGTCGTCCTTCCTCCCC 68

RESULT 15
AP003844      104481 bp  DNA      linear  HTG 04-JUL-2001
LOCUS      Oryza sativa chromosome 7 clone OJ1656_F06, *** SEQUENCING IN
DEFINITION  Oryza sativa chromosome 7 clone OJ1656_F06, *** SEQUENCING IN
ACCESSION  AP003844
VERSION    AP003844.1 GI:14595189
KEYWORDS  HTG; HTGS; PHASE2.
SOURCE
Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1656_F06.
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE
1 (bases 1 to 104481)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OJ1656_F06
JOURNAL
Published Only in Database (2001) In press
2 (bases 1 to 104481)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them

```

are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

- * NOTE: This is a 'working draft' sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

Location/Qualifiers

FEATURES

source

1. 104481
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="OJ165_F06"

BASE COUNT 28535 a 23142 c 23528 g 29276 t
ORIGIN

Query Match 84.4%; Score 22.8; DB 2; Length 104481;

Best Local Similarity 92.3%; Pred. No. 1e+02; Mismatches 24; Conservative 0; Indels 0; Gaps 0;

Oy 2 cccttcgtctgtctctctccct 27
|||||
Db 1269 CCCTTGTCGTCTCTCTCTCT 1294

Search completed: August 14, 2002, 21:49:18
Job time: 13576 sec


```

FT      /*tag- d
FT      /note= "putative Kozak consensus signal"
FT      258..323
FT      repeat_region
FT      /*tag- e
FT      /note= "encodes polyglutamine repeat region; contains
FT      repeats of CAG with 2 CAA codons interspersed"
FT      258..260
FT      repeat_unit
FT      /*tag- f
FT      /note= "CAG repeats"
FT      1..3986
FT      misc_feature
FT      /*tag- g
FT      /note= "sequence contained in DAN1 clone"
FT      3987..4200
FT      misc_feature
FT      /*tag- h
FT      /note= "derived from the EST's AAH92640, AAN90240 and
FT      AA213574 from dbEST database"
FT      4023..4029
FT      misc_feature
FT      /*tag- i
FT      /note= "region which differs in length between the
FT      sequences of the EST clones AAH92640, AAN90240
FT      and AA213574"
FT
FT      WO9717445-A1.
FT
FT      15-MAY-1997.
FT
FT      08-NOV-1996; 96WO-FR01773.
FT
FT      10-NOV-1995; 95FR-0013576.
FT
FT      (CNRS ) CNRS CENT NAT RECH SCI.
FT      (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
FT
FT      Lutz Y, Mandel J, Tora L, Trotter Y;
FT
FT      WPI: 1997-281034/25.
FT      P-PSDB; AAM24800, AAM24801.
FT
FT      Antibody 1C2 used for treating or preventing neuro-degenerative
FT      diseases - associated with proteins containing long poly:glutamine
FT      repeats, e.g. Huntington's disease
FT
FT      Claim 21: Page 45-47; 69pp; French.
FT
FT      The invention relates to a monoclonal antibody (Mab) 1C2 for the
FT      treatment of neurodegenerative diseases associated with the presence
FT      of polyglutamine repeat regions. This Mab is already known for its
FT      affinity to the TATA binding protein (TBP) transcription initiation
FT      factor, especially at the amino acid sequence LEEQOROOOQOQ found at
FT      the N-terminus of TBP. Mab 1C2 has been shown to have a high affinity
FT      for polyglutamine repeats with a proportional affinity to the number
FT      of glutamine repeats. This affinity has been used to identify genes
FT      encoding proteins containing long polyglutamine repeats which are
FT      implicated in neurodegenerative diseases. A screen of an expression
FT      library, generated from a lymphoblastic cell line from a patient
FT      suffering from spinocerebellar ataxia (SCA), with Mab 1C2 isolated 6
FT      new sequences (AA178906-T78911) encoding polyglutamine repeats. Mab 1C2
FT      also isolated the complete SCA2 gene in clone DAN1 (sequence presented
FT      here). The sequence appears to contain 2 open reading frames (ORF) the
FT      second of which may be generated by a frameshift slippage or by an
FT      alternative splicing event. The first ORF also encodes a 22 amino acid
FT      polyglutamine repeat region near the N-terminus of the protein. Normal
FT      SCA2 alleles contain 17-29 CAG triplet repeats with 1-3 CAA repeats
FT      interspersed whereas the mutant sequence from patients with SCA
FT      contains at least 30, preferably 37-50 CAG repeats.
FT      MAb 1C2, active fragment of it or nucleic acids encoding it are
FT      specifically used to treat Huntington's disease, SCA types 1-5 or 7,
FT      X-linked spinobulbar muscular atrophy (Kennedy disease),
FT      dentatorubral pallidolusial atrophy, dominant autosomal spinocerebellar
FT      ataxia, familial spastic paraplegia, bipolar affective disorder, manic
FT      depressive psychoses and schizophrenia.
FT
FT      Sequence 4200 BP; 1152 A; 1200 C; 913 G; 935 T; 0 other;

```

```

Query Match      100.0%; Score 27; DB 18; Length 4200;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 cccctcgctgcgtcctctccctcct 27
DB      68 cccctcgctgcgtcctctccctcct 94

RESULT 3
AAV30270
ID      AAV30270 standard; DNA; 4367 BP.
AC      AAV30270;
XX
XX      02-OCT-1998 (first entry)
XX
DE      Gene causative of spinocerebellar ataxia type 2 (SCA2) DNA sequence.
XX
KW      Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW      CAG repeat; neurodegenerative disease; ds.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      49..3990
FT      /*tag- a
FT      /product= "Spinocerebellar ataxia type 2 associated
FT      repeat_region 544..612
FT      /*tag- b
FT      /note= "normal CAG repeat region; this is increased in
FT      patients with SCA2"
FT      repeat_unit 544..546
FT      /*tag- c
XX
XX      WO9818920-A1.
XX
XX      07-MAY-1998.
XX
XX      30-OCT-1997; 97WO-JP03946.
XX
XX      30-OCT-1996; 96JP-0304059.
XX
XX      (SRLS-) SRL INC.
XX
XX      Sanpel K, Tsuji S;
XX
XX      WPI, 1998-272215/24.
XX      P-PSDB; AAM60213.
XX
XX      Nucleic acid fragments associated with spinocerebellar ataxia type 2
XX      - contain increased number of CAG repeat region compared to normal
XX      gene
XX
XX      Claim 1; Pages 13-22; 38pp; Japanese.
XX
XX      This represents the sequence of a gene causative of spinocerebellar
XX      ataxia type 2 (SCA2), a neurodegenerative disease. This gene associated
XX      with SCA2, has a tri-nucleotide (CAG) repeat region which in the
XX      expression product produces a polyglutamine sequence from Gln-166 to
XX      Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
XX      patients this number is increased to 35-100. Peptides encoded by nucleic
XX      acid fragments (DNA or RNA) containing sequences from the SCA2 associated
XX      gene, antibodies recognising the peptides and antisense nucleic acids
XX      hybridising with the nucleic acid fragments can be used for the
XX      investigation and diagnosis of SCA2. They can also be used for the
XX      treatment of SCA2 by antisense therapy or gene therapy.
XX
XX      Sequence 4367 BP; 1124 A; 1328 C; 991 G; 924 T; 0 other;

```

Query Match 100.0%; Score 27; DB 19; Length 4367;
 Best Local Similarity 100.0%; Pred. No. 0.33; Mismatches 0; Gaps 0;
 Matches 27; Conservative 0; Indels 0; Gaps 0;
 OY 1 ccccttcgtcgtccttcctccccc 27
 |||||||||||||||||||||||||
 Db 354 ccccttcgtcgtccttcctccccc 380

RESULT 4

AAV06552

ID AAV06552 standard; cDNA: 4481 BP.

XX AC AAV06552;

DT 06-JUL-1998 (first entry)

XX Human SCA2 cDNA including CAG repeat region.

XX SCA2 gene; spinocerebellar ataxia-2; ataxin-2; human;
 KM diagnosis: olivoponto-cerebellar atrophy; ss; ds.

XX OS Homo sapiens.

XX Key location/Qualifiers

FT CDS 164..4101

FT primer_bind complement (631..648)

FT primer_bind /note= "primer SCA2-A binding site"

FT primer_bind /tag= c

FT primer_bind /note= "primer SCA2-B binding site"

FT exon /tag= d

FT exon /note= "primer SCA2-14B binding site"

FT repeat_region /note= "predicted splice site"

FT repeat_unit /tag= f

FT repeat_unit /note= "CAG repeat region"

FT repeat_unit /tag= g

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= h

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= i

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= j

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= k

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= l

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= m

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= n

FT repeat_unit /note= "CAG repeat"

FT repeat_unit /tag= o

FT repeat_unit /note= "CAG repeat"

FT repeat_unit

691..693

FT repeat_unit

/tag= g

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

694..696

FT repeat_unit

/tag= i

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

700..702

FT repeat_unit

/tag= s

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

703..705

FT repeat_unit

/tag= t

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

706..708

FT repeat_unit

/tag= u

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

709..711

FT repeat_unit

/tag= v

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

712..714

FT repeat_unit

/tag= w

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

715..717

FT repeat_unit

/tag= x

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

718..720

FT repeat_unit

/tag= y

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

721..723

FT repeat_unit

/tag= z

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

721..723

FT repeat_unit

/tag= z

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

721..723

FT repeat_unit

/tag= z

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

721..723

FT repeat_unit

/tag= z

FT repeat_unit

/note= "CAG repeat"

FT repeat_unit

721..723

FT repeat_unit

/tag= z

FT repeat_unit

/note= "CAG repeat"

Claim 6: Page 52-58; 98pp; English.

Nucleic acids encoding human and mouse ataxin 2 - a product of the spinocerebellar ataxia 2 gene, SCA2; useful in the diagnosis of ataxia type 2

MP1: 1998-086523/08.
 P-PSDB; AAM33807.

(CEDA-) CEDARS SINAI MEDICAL CENT.
 Pulst S;

This cDNA sequence corresponds to a novel SCA2 gene encoding a human spinocerebellar ataxin-2 (SCA2) polypeptide, designated ataxin-2 (see AAM33807). A trisomy 21 foetal brain cDNA library and an adult human frontal cortex cDNA library in lambda Zapit were screened with probes obtained by PCR amplification of plasmid AAP5122B (see AAV06551). PCR products were used to screen the human adult frontal cortex library, and 5' clones were obtained by RT-PCR of placental mRNAs. Overlapping clones was used to generate the composite 4481 bp sequence. Ataxia type 2 can be diagnosed by detecting a genomic or transcribed mRNA sequence in an individual having an expanded CAG repeat at a location corresponding to the CAG repeat region of the SCA2 gene. The presence of at least 13 CAG repeats above the normal level (22, occasionally 23, repeats) is indicative of SCA2. Primers (see AAT99640-41) amplifying at least this region are used for diagnosis. Also claimed are kits for detecting mutations at the SCA2 locus, antisense oligonucleotides, and transgenic animals useful for studying the physiological roles of ataxin-2 and its effect upon behaviour.

SQ Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 27; DB 19; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
Db 468 cccctcgctgcgtccttcctccct 494

RESULT 5

AA23428
ID AA23428 standard; DNA; 4481 BP.

AC AA23428;

DT 19-JAN-2000 (first entry)

DE Human SCA2 DNA.

XX Proapoptotic; dependence domain; p75NTR; androgen receptor; DCC;
KW huntingtin polypeptide; Machado-Joseph disease; SCA1; SCA2; SCA6;
KW atrophin-1; cell death; apoptosis; Huntington's disease; head trauma;
KW Alzheimer's disease; Kennedy's disease; spinocerebellar ataxia; stroke;
KW dentatorubropallidoluysian atrophy; cell proliferation; cell survival;
KW neoplastic; malignant; autoimmune; fibrotic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 163..4101

FT /tag= a

FT /product= "SCA2"

XX MO9945944-A1.

XX 16-SEP-1999.

XX PF 11-MAR-1999; 99WO-US05250.

XX PR 12-MAR-1998; 98US-0041886.

XX (BURN-) BURNHAM INST.

XX Bredesen DE, Rabizadeh S;

XX WPI: 1999-561617/47.

XX P-PSDB; AA133495.

XX New proapoptotic dependence peptides, used to develop products for

XX treating, e.g. Alzheimer's disease -

XX Disclosure; Page 130-135; 199pp; English.

XX This invention describes novel pure proapoptotic dependence peptides,
CC which comprise a sequence of an active dependence domain selected from
CC dependence polypeptides consisting of p75NTR, androgen receptor, DCC,
CC huntingtin polypeptide, Machado-Joseph disease gene product, SCA1, SCA2,
CC SCA6 and atrophin-1 polypeptide. The proapoptotic peptides are capable
CC of inducing cell death and can be used to develop products to mediate or
CC inhibit apoptosis. The methods can be used for reducing the severity of
CC a proapoptotic dependence domain mediated pathological conditions e.g.

CC Huntington's disease, Alzheimer's disease, Kennedy's disease,

CC Spinocerebellar ataxias, dentatorubropallidoluysian atrophy,

CC Machado-Joseph disease, stroke or head trauma. They can also be used for

CC reducing the severity of a pathological condition mediated by upregulated

CC cell proliferation or cell survival e.g. neoplastic, malignant,

CC autoimmune or fibrotic conditions. This sequence encodes the human

CC SCA2 polypeptide described in the method of the invention.

XX Sequence 4481 BP; 1144 A; 1380 C; 1014 G; 943 T; 0 other;

Query Match 100.0%; Score 27; DB 20; Length 4481;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
Db 468 cccctcgctgcgtccttcctccct 494

RESULT 6

AAV17224
ID AAV17224 standard; DNA; 355 BP.

AC AAV17224;

DT 29-JUN-1998 (first entry)

DE SCA2 gene fragment.

XX SCA2 gene; spinocerebellar ataxia type II; CAG repeat; PCR primer; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 341..355

FT /tag= a

FT /note= "SCA2 protein fragment"

XX MO9803679-A1.

XX 29-JAN-1998.

XX PF 18-JUL-1996; 96WO-JP01999.

XX PR 18-JUL-1996; 96WO-JP01999.

XX (SRLS-) SRL INC.

XX Sanpei K, Tsuji S;

XX WPI: 1998-120796/11.

XX P-PSDB; AAV41370.

XX Diagnosing spinocerebellar ataxia type II - by PCR and determining

XX number of CAG repeat units

XX Claim 1; Page 10; 23pp; Japanese.

XX This sequence represents a fragment of the SCA2 gene. It can be used in

XX the method of the invention for diagnosing spinocerebellar ataxia type

XX II, by performing PCR on the test DNA using two primers hybridizing to

XX parts of the SCA2 gene sequence, and determining the number of CAG

XX repeats in the amplified products. The method provides an easy means for

XX the diagnosis of spinocerebellar ataxia type II.

XX Sequence 355 BP; 20 A; 176 C; 102 G; 55 T; 2 other;

Query Match 98.5%; Score 26.6; DB 19; Length 355;
Best Local Similarity 96.3%; Pred. No. 0.44;
Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
Db 166 cccctcgctgcgtccttcctccct 192

RESULT 7

AAV17229
ID AAV17229 standard; DNA; 623 BP.

XX

```

AC  AAV17229;
XX
DT  29-JUN-1998 (first entry)
XX
XX  SCA2 gene fragment.
XX
XX  SCA2 gene; spinocerebellar ataxis type II; CAG repeat; PCR primer; ss.
XX
XX  Synthetic.
XX
XX  Key          Location/Qualifiers
XX  CDS          341..583
XX               /*tag= a
XX               /note= "SCA2 protein fragment, no stop codon given"
XX
XX  WO9803679-A1.
XX
XX  29-JAN-1998.
XX
XX  18-JUL-1996; 96WO-JP01999.
XX
XX  18-JUL-1996; 96WO-JP01999.
XX
XX  (SRLS-) SRL INC.
XX
XX  Sanpel K, Tsuji S;
XX
XX  WPI: 1998-120796/11.
XX
XX  P-PSDB; AAM41372.
XX
XX  Diagnosing spinocerebellar ataxis type II - by PCR and determining
XX  number of CAG repeat units
XX
XX  Example 1; Page 11-12; 23pp; Japanese.
XX
XX  This sequence represents a fragment of the SCA2 gene. It can be used in
XX  the method of the invention for diagnosing spinocerebellar ataxis type
XX  II, by performing PCR on the test DNA using two primers hybridizing to
XX  parts of the SCA2 gene sequence, and determining the number of CAG
XX  repeats in the amplified products. The method provides an easy means for
XX  the diagnosis of spinocerebellar ataxis type II.
XX
XX  Sequence 623 BP; 55 A; 292 C; 189 G; 85 T; 2 other;
XX
XX  Query Match          98.5%; Score 26.6; DB 19; Length 623;
XX  Best Local Similarity 96.3%; Pred. No. 0.45;
XX  Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 ccccttcgctgcgtcctctccct 27
XX  ||||| ||||| ||||| ||||| |||||
XX  Db 166 ccccttcgctgcgtcctctccct 192
XX
XX  RESULT 8
XX  AAS80334
XX  ID AAS80334 standard; cDNA; 419 BP.
XX
XX  AAS80334;
XX
XX  13-FEB-2002 (first entry)
XX
XX  DNA encoding novel human diagnostic protein #16138.
XX
XX  Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX  food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX  Homo sapiens.
XX
XX  WO200175067-A2.
XX
XX  11-OCT-2001.
XX

```

```

PF  30-MAR-2001; 2001WO-US08631.
XX
XX  31-MAR-2000; 2000US-0540217.
XX  23-AUG-2000; 2000US-0649167.
XX
XX  (HXSE-) HXSEQ INC.
XX
XX  Drmanac RT, Liu C, Tang YT;
XX
XX  WPI: 2001-639362/73.
XX  P-PSDB; ABG16147.
XX
XX  New isolated polynucleotide and encoded polypeptides, useful in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits and to assess
XX  biodiversity -
XX
XX  Claim 1; SEQ ID NO 16138; 103pp; English.
XX
XX  The invention relates to isolated polynucleotide (I) and
XX  polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX  and gene mapping, and in recombinant production of (II). The
XX  polynucleotides are also used in diagnostics as expressed sequence tags
XX  for identifying expressed genes. (I) is useful in gene therapy techniques
XX  to restore normal activity of (II) or to treat disease states involving
XX  (II). (II) is useful for generating antibodies against it, detecting or
XX  quantitating a polypeptide in tissue, as molecular weight markers and as
XX  a food supplement. (II) and its binding partners are useful in medical
XX  imaging of sites expressing (II). (I) and (II) are useful for treating
XX  disorders involving aberrant protein expression or biological activity.
XX  The polypeptide and polynucleotide sequences have applications in
XX  diagnostics, forensics, gene mapping, identification of mutations
XX  responsible for genetic disorders or other traits to assess biodiversity
XX  and to produce other types of data and products dependent on DNA and
XX  amino acid sequences. AAS64197-AAS94564 represent novel human
XX  diagnostic coding sequences of the invention.
XX  Note: The sequence data for this patent did not appear in the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 419 BP; 87 A; 122 C; 106 G; 104 T; 0 other;
XX
XX  Query Match          82.2%; Score 22.2; DB 23; Length 419;
XX  Best Local Similarity 88.9%; Pred. No. 18;
XX  Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX  1 ccccttcgctgcgtcctctccct 27
XX  ||||| ||||| ||||| ||||| |||||
XX  Db 183 ccccttcgctgcgtcctctccct 209
XX
XX  RESULT 9
XX  ABL26791/c
XX  ID ABL26791 standard; DNA; 2181 BP.
XX
XX  ABL26791;
XX
XX  26-MAR-2002 (first entry)
XX
XX  Drosophila melanogaster genomic polynucleotide SEQ ID NO 31846.
XX
XX  Drosophila; developmental biology; cell signalling; insecticide;
XX  pharmaceutical; gene; ds.
XX
XX  Drosophila melanogaster.
XX
XX  WO200171042-A2.
XX
XX  27-SEP-2001.
XX
XX  23-MAR-2001; 2001WO-US09231.
XX

```

```

XX 23-MAR-2000: 2000US-191637P.
PR 11-JUL-2000: 2000US-0614150.
XX (PEKE ) PE CORP NX.
XX Venter JC, Adams M, Li PMD, Myers EW;
XX WPI: 2001-656660/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 31846; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX sequences (AB101840-AB1616175) and the encoded proteins
XX (AB85737-AB872072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at flp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2181 BP; 503 A; 685 C; 650 G; 343 T; 0 other:
SO

```

Query Match 76.3%; Score 20.6; DB 23; Length 2181;
 Best Local Similarity 85.2%; Pred. No. 71;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 cccctcgctgctctctccct 27
DB 1951 CCCCTGCTGCTGCTCTCTCTCTCACT 1925

```

RESULT 10
 AAQ37948/C
 ID AAQ37948 standard; cDNA; 2353 BP.
 XX
 AC AAQ37948;
 XX
 DT 11-JUL-1993 (first entry)
 XX
 DE Sequence of a DNA isolate encoding eukaryotic transcription factor
 (TF).
 XX
 KW Transcription factor: Y1; modulatory adeno-associated virus type 2;
 KW Epstein-Barr virus; oncogene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT polyA_signal 1494..1498
 FT polyA_signal /*tag- a
 FT polyA_signal 2333..2338 /*tag- b
 FT CDS 241..1485 /*tag- c
 FT
 XX WO9304076-A.
 XX
 XX 04-MAR-1993.
 XX
 XX 14-AUG-1992; 92MO-US06840.
 XX
 XX 16-AUG-1991; 91US-0746485.
 XX
 XX (UYPR-) UNIV PRINCETON.

```

XX Seto E, Shenk T, Shi Y;
XX WPI: 1993-093929/11.
XX P-PSDB; AAR32020.
XX
XX purified and isolated mammalian transcription factor Y1 - used
XX to modulate transcription in adeno-associated virus type 2-
XX Epstein-Barr viruses and oncogenes; useful in diagnosis and
XX screening
XX
XX Claim 32; Fig 11; 100pp; English.
XX
XX The cDNA of transcription factor (TF) Y1 (AAQ37948) was derived from
XX HeLa cells derived from cervical carcinoma from clone p14-1 or pY1
XX of the D98/AH-2 library. TF Y1 has a mol. wt. of 68 kD and is
XX capable of binding to a sequence between -50 and -70 (PS-60 site)
XX and to the transcription initiation region of the promoter (PS+1
XX site) of an adeno-associated virus (AAV). It represses transcription
XX directed by a TATA element plus initiator (Inr) sequence. It can
XX mediate transcription through an Inr sequence or element. The
XX initiation site is of a lymphocyte-specific terminal
XX deoxynucleotidyltransferase gene (TdtInr) or human leucocyte IFN gene
XX (leIF-3 Inr). Y1 can effect latency of viruses. It is a zinc
XX finger protein and has a glycine rich sequence and an acidic
XX domain.
XX
XX Sequence 2353 BP; 627 A; 598 C; 607 G; 521 T; 0 other:
SO

```

Query Match 76.3%; Score 20.6; DB 14; Length 2353;
 Best Local Similarity 85.2%; Pred. No. 71;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY 1 cccctcgctgctctctccct 27
DB 397 CCTGCTGCTGCTGCTCTCTCTCTCT 371

```

RESULT 11
 AAV07530/C
 ID AAV07530 standard; cDNA; 2353 BP.
 XX
 AC AAV07530;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Human transcription factor Y1 cDNA.
 XX
 KW Y1; transcription factor; human immunodeficiency virus; HIV; AIDS;
 KW infection; retrovirus; therapy; LSF; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 241..1485 /*tag- a
 FT
 XX WO9833067-A1.
 XX
 XX 30-JUL-1998.
 XX
 XX 13-JAN-1998; 98MO-US00574.
 XX
 XX 23-JAN-1997; 97US-0036242.
 XX
 XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
 XX
 XX Devico A, Margolis D, Romero F;
 XX WPI: 1998-428099/36.
 XX P-PSDB; AAW65406.
 XX

PT Transcription factors YY1 and LSF, and their derivative(s) and
PT analogue(s) - useful for, e.g. Inhibiting HIV transcription,
PT replication and/or infection in vitro or in vivo or preventing
PT disorders associated with HIV infection
XX
PS Disclosure: Page 74-76; 112pp; English.
XX
CC This is the nucleotide sequence of cDNA encoding human YY1 (see
CC AAM65406), a multifunctional transcription factor that has previously
CC been shown to bind to the HIV-1 LTR (see AAV07529) and to repress
CC HIV-1 transcription and virion production. The invention relates
CC to compositions of YY1 and LSF (see AAM65407) (including peptides),
CC or nucleic acids encoding YY1, LSF or their peptides, that are
CC effective at inhibiting HIV transcription, replication and/or
CC infection in vitro or in vivo, decreasing viral load, and/or
CC treating or preventing disorders associated with HIV infection.
CC The invention also provides for treatment of HIV infection by
CC administration of an inhibitor of the complex comprising YY1 and
CC LSF (i.e. the RGS complex) thereby inhibiting the repression of HIV
CC transcription and releasing latent HIV from its state of latency.
CC Upon release of the virus from latency, antiviral agents are
CC administered to clear the infection.
XX
SQ Sequence 2353 BP; 625 A; 600 C; 606 G; 522 T; 0 other;

Query Match 76.3%; Score 20.6; DB 19; Length 2353;
Best Local Similarity 85.2%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ccccttcgtcgtcgtccttcctccct 27
||| ||||| ||||| ||||| |||||
DB 397 CCTCGTGTGTGTGTCTCTCTCTCTCT 371

RESULT 12
AAH02898/C
ID AAH02898 standard; DNA: 2353 BP.
XX
AC AAH02898;
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 49.
XX
KW Human; shear stress-response protein; vascular disease;
KW arteriosclerosis; ds.
XX
OS Homo sapiens.
XX
PN WO200125427-A1.
XX
PD 12-APR-2001.
XX
PF 02-OCT-2000; 2000WO-JP06840.
XX
PR 01-OCT-1999; 99JP-0280976.
XX
PA (KYOM) KYOMA HAKKO KOGYO KK.
PA (NOJI/) NOJIMA H.
XX
PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
DR WPI: 2001-266308/27.
DR P-PSDB: AAB90775.
XX
PT DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
PS Claim 20; Page 356-360; 678pp; Japanese.
XX

CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 2353 BP; 625 A; 598 C; 608 G; 522 T; 0 other;

Query Match 76.3%; Score 20.6; DB 22; Length 2353;
Best Local Similarity 85.2%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ccccttcgtcgtcgtccttcctccct 27
||| ||||| ||||| ||||| |||||
DB 397 CCTCGTGTGTGTGTCTCTCTCTCTCT 371

RESULT 13
ABL26790
ID ABL26790 standard; DNA: 4995 BP.
XX
AC ABL26790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 31843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AAB57737-AAB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4995 BP; 1150 A; 1198 C; 1235 G; 1412 T; 0 other;

Query Match 76.3%; Score 20.6; DB 23; Length 4995;
Best Local Similarity 85.2%; Pred. No. 72;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtctctccct 27
DB 1231 cccctcgctgcgtctctccact 1257

RESULT 14
ID AAS26629/c
AAS26629 standard; DNA; 21724 BP.

AC AAS26629;
XX
XX
DT 07-NOV-2001 (first entry)
XX
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1603.

XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;
XX cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnary; secreted protein; rheumatoid arthritis;
XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX corneal infection; wound healing; epithelial cell proliferation;
XX skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.
XX
XX PN MO200155322-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001MO-US01341.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
XX
PS Disclosure: SEQ ID No 1603; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence is a genomic DNA encoding a partial novel secreted protein of
CC the invention.

Query Match 76.3%; Score 20.6; DB 22; Length 21724;
Best Local Similarity 85.2%; Pred. No. 73;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cccctcgtcgtcgtccctccct 27
||||| ||||| ||||| |||||
Db 374 ccccttcttctgtctctctctctctct 348

RESULT 15
AAK86125/c

ID AAK86125 standard; DNA; 21724 BP.
XX
XX AAK86125;
AC
XX
XX 07-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40937.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 07-JUN-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 30-JUN-2000; 2000US-0214886.
PR 07-JUL-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.

PR	14-SEP-2000	2000US-02332399
PR	14-SEP-2000	2000US-02332401
PR	14-SEP-2000	2000US-02332410
PR	14-SEP-2000	2000US-02333663
PR	14-SEP-2000	2000US-02333664
PR	14-SEP-2000	2000US-02333665
PR	14-SEP-2000	2000US-02333666
PR	21-SEP-2000	2000US-02344723
PR	21-SEP-2000	2000US-02344724
PR	23-SEP-2000	2000US-02343977
PR	23-SEP-2000	2000US-02343978
PR	26-SEP-2000	2000US-02354844
PR	27-SEP-2000	2000US-02358534
PR	27-SEP-2000	2000US-02358536
PR	29-SEP-2000	2000US-02365277
PR	29-SEP-2000	2000US-02365278
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370340
PR	13-OCT-2000	2000US-02379335
PR	13-OCT-2000	2000US-02393937
PR	20-OCT-2000	2000US-02409360
PR	20-OCT-2000	2000US-02411821
PR	20-OCT-2000	2000US-02411825
PR	20-OCT-2000	2000US-02417866
PR	20-OCT-2000	2000US-02417867
PR	20-OCT-2000	2000US-02418078
PR	20-OCT-2000	2000US-02418109
PR	20-OCT-2000	2000US-02418266
PR	01-NOV-2000	2000US-02446177
PR	08-NOV-2000	2000US-02464775
PR	08-NOV-2000	2000US-02464776
PR	08-NOV-2000	2000US-02464777
PR	08-NOV-2000	2000US-02465127
PR	08-NOV-2000	2000US-02465128
PR	08-NOV-2000	2000US-02465132
PR	08-NOV-2000	2000US-02465623
PR	08-NOV-2000	2000US-02465629
PR	08-NOV-2000	2000US-02466110
PR	08-NOV-2000	2000US-02466111
PR	08-NOV-2000	2000US-02466113
PR	17-NOV-2000	2000US-02492507
PR	17-NOV-2000	2000US-02492508
PR	17-NOV-2000	2000US-02492514
PR	17-NOV-2000	2000US-02492515
PR	17-NOV-2000	2000US-02492516
PR	17-NOV-2000	2000US-02492517
PR	17-NOV-2000	2000US-02492910
PR	17-NOV-2000	2000US-02492911
PR	17-NOV-2000	2000US-02492912
PR	17-NOV-2000	2000US-02492913
PR	17-NOV-2000	2000US-02492914
PR	17-NOV-2000	2000US-02492915
PR	17-NOV-2000	2000US-02492916
PR	17-NOV-2000	2000US-02492917
PR	17-NOV-2000	2000US-02492918
PR	17-NOV-2000	2000US-02492919
PR	17-NOV-2000	2000US-02492920
PR	17-NOV-2000	2000US-02501660
PR	01-DEC-2000	2000US-02503591
PR	05-DEC-2000	2000US-02510388
PR	05-DEC-2000	2000US-02510388

PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251889.
PR	08-DEC-2000;	2000US-0251907.
PR	11-DEC-2000;	2000US-0251490.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM,	
XX	WPI: 2001-483426/52.	

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

PS Disclosure; SEQ ID NO 40937; 3071pp + Sequence Listing; English.

CC AAM54950 AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAM87654 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAM54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 21724 BP: 6980 A: 4481 C:4432 G: 5831 T: 0 other:

	Query Match	76.3%	Score 20.6	DB 22	Length 21724
	Best Local Similarity	85.2%	Pred. No. 73		
	Matches 23	Conservative	0	Mismatches 4	Indels 0
QY	1	ccccctcgtcgtcgtcctctcccccct	27		
db	374	cccccttcgtcgtcgtcctctcccccct	348		

Search completed: August 14, 2002, 22:06:56
Job time: 11711 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:57:32 ; Search time 203.42 Seconds
(without alignments)
32.603 Million cell updates/sec

Title: US-09-707-919-10

Perfect score: 27
Sequence: 1 ccccttcgtcgtcgtccttcctccccc 27

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	4481	4	US-09-041-886-18 Sequence 18, Appl
2	26.6	98.5	355	4	US-09-043-303-1 Sequence 1, Appl
3	26.6	98.5	623	4	US-09-043-303-5 Sequence 5, Appl
4	20.6	76.3	2353	4	PCR-US92-06840-1 Sequence 1, Appl
5	19.2	71.1	572	4	US-08-998-416-716 Sequence 716, App
6	19.2	71.1	2277	1	US-08-676-967-5 Sequence 5, Appl
7	19.2	71.1	2277	1	US-08-676-974-5 Sequence 5, Appl
8	19.2	71.1	2277	2	US-09-098-487-5 Sequence 5, Appl
9	19.2	71.1	2523	2	US-08-410-784A-3 Sequence 3, Appl
10	19.2	71.1	2712	2	US-08-410-784A-1 Sequence 4, Appl
11	19	70.4	561	1	US-08-832-883-4 Sequence 4, Appl
12	19	70.4	561	2	US-08-832-877-4 Sequence 4, Appl
13	19	70.4	1518	4	US-09-257-581-4 Sequence 4, Appl
14	19	70.4	1518	4	US-09-257-581-6 Sequence 6, Appl
15	19	70.4	1860	5	US-08-331-644-3 Sequence 3, Appl
16	19	70.4	1860	5	PCR-US93-04102-3 Sequence 3, Appl
17	19	70.4	2461	1	US-08-832-883-3 Sequence 3, Appl
18	19	70.4	2461	2	US-08-832-877-113 Sequence 113, App
19	19	70.4	4853	1	US-08-832-883-1 Sequence 1, Appl
20	19	70.4	4853	2	US-08-832-877-1 Sequence 1, Appl
21	19	70.4	5035	2	US-08-882-083-1 Sequence 1, Appl
22	19	70.4	5035	3	US-08-558-107-1 Sequence 1, Appl
23	19	70.4	5035	3	US-09-243-539-1 Sequence 1, Appl
24	19	70.4	9837	1	US-08-832-883-68 Sequence 68, Appl
25	19	70.4	9837	2	US-08-832-877-68 Sequence 68, Appl
26	18.8	69.6	1554	1	US-08-463-115-3 Sequence 3, Appl
27	18.8	69.6	1554	1	US-08-465-388-3 Sequence 3, Appl

28	18.6	68.9	11703	4	US-09-101-886B-3 Sequence 3, Appl
29	18.2	67.4	2384	1	US-07-814-964-10 Sequence 10, Appl
30	18.2	67.4	2384	1	US-08-258-442-10 Sequence 10, Appl
31	18.2	67.4	2384	1	US-08-328-809-5 Sequence 5, Appl
32	18.2	67.4	2384	5	PCR-US92-11107-10 Sequence 10, Appl
33	18.2	67.4	3133	4	US-09-193-069-1 Sequence 1, Appl
34	17.6	65.2	48	4	US-08-979-608A-36 Sequence 36, Appl
35	17.6	65.2	84	4	US-08-979-608A-37 Sequence 37, Appl
36	17.6	65.2	189	2	US-08-733-505A-51 Sequence 51, Appl
37	17.6	65.2	189	2	US-08-733-505A-52 Sequence 52, Appl
38	17.6	65.2	189	2	US-08-733-505A-53 Sequence 53, Appl
39	17.6	65.2	189	2	US-08-733-505A-54 Sequence 54, Appl
40	17.6	65.2	715	4	US-09-247-155-139 Sequence 139, App
41	17.6	65.2	944	1	US-08-665-617-1 Sequence 1, Appl
42	17.6	65.2	946	2	US-08-717-123-1 Sequence 1, Appl
43	17.6	65.2	966	2	US-08-766-738-2 Sequence 2, Appl
44	17.6	65.2	1105	3	US-08-985-335-2 Sequence 2, Appl
45	17.6	65.2	1105	4	US-09-410-372-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-18
Sequence 18, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 4481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 163..4099
US-09-041-886-18

Query Match 100.0%; Score 27; DB 4; Length 4481;
Best local Similarity 100.0%; Pred. No. 0.067;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
DB 468 cccctcgctgcgtccttcctccct 494

RESULT 2
US-09-043-303-1
: Sequence 1, Application US/09043303
: Patent No. 6251589
: GENERAL INFORMATION:
: APPLICANT: TSUJI, Shoji
: APPLICANT: SANPEI, Kazuhiro
: TITLE OF INVENTION: Method for Diagnosing SpinoCerebellar Ataxia Type 2 and
: TITLE OF INVENTION: Primers Therefor
: FILE REFERENCE: 0760-0241P
: CURRENT APPLICATION NUMBER: US/09/043,303
: CURRENT FILING DATE: 1998-05-18
: EARLIER APPLICATION NUMBER: PCT/JP96/01999
: EARLIER FILING DATE: 1996-07-18
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 355
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (341)..(355)
US-09-043-303-1

Query Match 98.5%; Score 26.6; DB 4; Length 355;
Best Local Similarity 96.3%; Pred. No. 0.09;
Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccctcgctgcgtccttcctccct 27
|||||
DB 166 cccctcgctgcgtccttcctccct 192

RESULT 3
US-09-043-303-5
: Sequence 5, Application US/09043303
: Patent No. 6251589
: GENERAL INFORMATION:
: APPLICANT: TSUJI, Shoji
: APPLICANT: SANPEI, Kazuhiro
: TITLE OF INVENTION: Method for Diagnosing SpinoCerebellar Ataxia Type 2 and
: TITLE OF INVENTION: Primers Therefor
: FILE REFERENCE: 0760-0241P
: CURRENT APPLICATION NUMBER: US/09/043,303
: CURRENT FILING DATE: 1998-05-18
: EARLIER APPLICATION NUMBER: PCT/JP96/01999
: EARLIER FILING DATE: 1996-07-18
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 623
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (341)..(583)
: FEATURE:
: OTHER INFORMATION: Tsp-2
US-09-043-303-5

Query Match 98.5%; Score 26.6; DB 4; Length 623;
Best Local Similarity 96.3%; Pred. No. 0.091;
Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 cccctcgctgcgtccttcctccct 27

DB 166 cccctcgctgcgtccttcctccct 192
|||||

RESULT 4
PCT-US92-06840-1/c
: Sequence 1, Application PC/TUS9206840
: GENERAL INFORMATION:
: APPLICANT: Shi, Yang
: APPLICANT: Seto, Edward
: APPLICANT: Shenk, Thomas
: TITLE OF INVENTION: YY1 TRANSCRIPTION FACTOR AND METHODS OF
: TITLE OF INVENTION: ISOLATING SAME
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
: STREET: 1180 Avenue of the Americas - 7th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-8403
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/06840
: FILING DATE: 19920814
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/746,485
: FILING DATE: 16-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Dennis, Manette
: REGISTRATION NUMBER: 30,623
: REFERENCE/DOCKET NUMBER: M-12594 CIP (1570-8)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 382-0700
: TELEFAX: (212) 382-0888
: TELEX: 236925
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2353 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: TISSUE TYPE: Hela cells derived from cervical
: TISSUE TYPE: carcinoma
: CELL TYPE: tumor cells
: CELL LINE: Hela
: IMMEDIATE SOURCE:
: LIBRARY: D98/AH-2
: CLONE: p14-1 or pYY1
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 241..1485
PCT-US92-06840-1

Query Match 76.3%; Score 20.6; DB 5; Length 2353;
Best Local Similarity 85.2%; Pred. No. 13;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 cccctcgctgcgtccttcctccct 27
|||
DB 397 CCTGCTGCTGCTGCTGCTGCTGCT 371

```
RESULT 5
US-08-998-416-716
; Sequence 716, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippson, Peter
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38, 241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 716:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 572 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1469UP
US-08-998-416-716

Query Match 71.1%; Score 19.2; DB 4; Length 572;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 ctcgcgcgcgcctctccctc 27
DB 141 cttcgctgcgcctctctcct 164

RESULT 6
US-08-676-967-5/c
; Sequence 5, Application US/0867967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; MOLECULE TYPE: CDNA
US-08-676-967-5

Query Match 71.1%; Score 19.2; DB 1; Length 227;
Best Local Similarity 87.5%; Pred. No. 42;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 ctcgcgcgcgcctctccctc 27
DB 712 CGTGCCTGCTGCTCTCTCTCT 689

RESULT 7
US-08-676-974-5/c
; Sequence 5, Application US/0867974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
```

```

? SEQUENCE CHARACTERISTICS: ?
? LENGTH: 2277 base pairs ?
? TYPE: nucleic acid ?
? STRANDEDNESS: double ?
? TOPOLOGY: linear ?
? MOLECULE TYPE: CDNA ?
? US-08-676-974-5

```

Query Match	71.1%;	Score 19.2;	DB 1;	Length 2277;
Best Local Similarity	87.5%;	Pred. No. 42;		
Matches 21; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

```

RESULT 8
US-09-098-487-5/c
Sequence 5, Application US/09098487
Patent No. 5917025
GENERAL INFORMATION:
APPLICANT: COLLINS, Kathleen
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,487
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-098-487-5

```

Query Match	71.1%	Score 19.2	DB 2	Length 2277
Best Local Similarity	87.5%	Pred. No. 42		
Matches 21	Conservative 0	Mismatches 3	Indels 0	Gaps 0
OY	4	ctctgctgctctctctctccct	27	
b	712	CGTCGCTCGTCGTCGTCCTC	689	

RESULT 9
US-08-410-784A-3/c
; Sequence 3, Application US/08410784A
; Patent No. 5912413
; GENERAL INFORMATION:

APPLICANT: MYERS, ALAN M.
 APPLICANT: JAMES, MARTHA G.
 TITLE OF INVENTION: ISOLATION OF SU1, A STARCH DEBRANCHING
 TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
 TITLE OF INVENTION: SUGARY 1
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP
 STREET: Ten Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/410,784A
 FILING DATE: 24-MAR-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Helme, Ph.D., Holliday C
 REGISTRATION NUMBER: 34,346
 REFERENCE/DOCKET NUMBER: ISU-002XX
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-542-2290
 TELEFAX: 617-451-0313
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2523 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 US-08-410-784A-3

```

Query Match      71.1%  Score 19.2;  DB 2;  Length 2523;
Best Local Similarity 87.5%;  Pred No. 42;
Matches 21;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0

```

RESULT 10784A-1/c
 US-08-410-784A-1/c
 : Sequence 1, Application US/08410784A
 : Patent No. 5912413
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: MYERS, ALAN M.
 : APPLICANT: JAMES, MARTHA G.
 : TITLE OF INVENTION: ISOLATION OF SU1, A STARCH DEBRANCHING
 : TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE
 : TITLE OF INVENTION: SUGARY 1
 : NUMBER OF SEQUENCES: 9
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Weingarten, Schurgin, Gagnebin and Hayes LLP
 : STREET: Ten Post Office Square
 : CITY: Boston
 : STATE: MA
 : COUNTRY: USA

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2002, 21:04:43 ; Search time 7749.14 Seconds

(without alignments)
47.027 Million cell updates/sec

Title: US-09-707-919-10

Sequence: 1 cccctgcgcgcgcctctcccccct 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pla:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	482	9	AL039573 DKFZp434D1311
2	27	100.0	500	10	BI547486 603191091
3	25.4	94.1	1100	10	BM455214 ACENECOURT
4	24.4	90.4	126	10	F14808 SSC20D02 Po
5	22.8	84.4	385	12	AO911478 LMAJFV1.1
6	22.2	82.2	192	10	BE814445 MRO-BM007
7	22.2	82.2	472	10	BE515471 UT-H-BM1-
8	22.2	82.2	473	9	AI423306 LF50905.x
9	22.2	82.2	548	9	AM452627
10	22.2	82.2	680	10	BF612947
11	22.2	82.2	1201	12	CNS0164L
12	21.4	79.3	513	10	BM324144 P1C1_24.F
13	21.2	78.5	538	10	BF277889 GA_EB003
14	21.2	78.5	575	10	BF616867 HVSMEC001
15	21.2	78.5	629	10	BG300411 HVSMEB001
16	21.2	78.5	629	10	BG300416 HVSMEB001
17	21.2	78.5	767	10	BI454901 603173450

C 18	21.2	78.5	782	10	BF617725 HVSMEC001
C 19	21.2	78.5	998	12	AL304316 Tetradon
C 20	20.8	77.0	373	12	A2270630 RPT-23-4
C 21	20.8	77.0	399	9	BE130057 945035H09
C 22	20.8	77.0	403	9	AM676976 DC1.3.G08
C 23	20.8	77.0	425	10	BE238538 946003H06
C 24	20.8	77.0	428	9	AM287536 LG1_242_C
C 25	20.8	77.0	436	10	BI778622 EBFO07_SQ
C 26	20.8	77.0	470	9	BE186337 945040A08
C 27	20.8	77.0	522	10	BM325922 P1C1_54.E
C 28	20.8	77.0	540	10	BE123362 945040A08
C 29	20.8	77.0	540	10	BF728745 1000066P0
C 30	20.8	77.0	542	9	AM787811 945003H06
C 31	20.8	77.0	545	9	BE025241 945025B05
C 32	20.8	77.0	562	9	AM927813 945003H06
C 33	20.8	77.0	590	9	AM455692 707089G01
C 34	20.8	77.0	655	10	BM442503 EBAN01_SQ
C 35	20.8	77.0	706	10	BM442231 EBAN01_SQ
C 36	20.8	77.0	828	10	BG418286 HVSMEC002
C 37	20.8	77.0	1192	10	BF965843 602277484
C 38	20.6	76.3	135	10	BE575131 946087D09
C 39	20.6	76.3	293	9	AU082338
C 40	20.6	76.3	298	10	BI307056 PRSS0361
C 41	20.6	76.3	299	9	AM422960 f166D09.y
C 42	20.6	76.3	332	10	C71859 C71859
C 43	20.6	76.3	344	10	BE552922 946087D09
C 44	20.6	76.3	360	10	BI549384 603190039
C 45	20.6	76.3	362	9	AM422592 f144C08.y

ALIGNMENTS

RESULT 1
LOCUS AL039573 482 bp mRNA linear EST 29-FEB-2000
DEFINITION DKFZp434D1311.t1 434 (synonym: htes3) Homo sapiens CDNA clone
ACCESSION AL039573
VERSION AL039573.1 GI:5408612
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 482)
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
S.
EST (Duesterhoeft, et al.)
TITLE Unpublished (1999)
JOURNAL Contact: Duesterhoeft A
COMMENT MIPS

Am Klopferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by QIAGEN (Hilden/Germany) within the CDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp434D1311) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

FEATURES

source

1. 482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434D1311"
/clone_id="434 (synonym: htes3)"
/issue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
49 a 218 c 145 g 70 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 27; DB 9; Length 482;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccccttgctgcgtcctcttcctccct 27
|||||
Db 115 CCCCTTCGTCGTCCTTCCTCCCT 141

RESULT 2
LOCUS B1547486 500 bp mRNA linear EST 05-SEP-2001
DEFINITION 603191091P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262335 5',
mRNA sequence.
ACCESSION B1547486
VERSION B1547486.1 GI:15434798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 500)
NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11661 row: e column: 24
High quality sequence stop: 485.
Location/Qualifiers
1. 500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262335"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (glucag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 57 a 222 c 150 g 71 t
ORIGIN

Query Match 100.0%; Score 27; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ccccttgctgcgtcctcttcctccct 27
|||||
Db 118 CCCCTTCGTCGTCCTTCCTCCCT 144

RESULT 3
LOCUS B1545214 1100 bp mRNA linear EST 05-FEB-2002

DEFINITION AGENCOURT_6405612 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5500163
5', mRNA sequence.
ACCESSION B1545214
VERSION B1545214.1 GI:18504254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1100)
NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12134 row: k column: 12
High quality sequence stop: 623.
Location/Qualifiers
1. 1100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5500163"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 240 a 329 c 306 g 219 t 6 others
ORIGIN

Query Match 94.1%; Score 25.4; DB 10; Length 1100;
Best Local Similarity 96.3%; Pred. No. 2e+02;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ccccttgctgcgtcctcttcctccct 27
|||||
Db 89 CCCCTTCGTCGTCCTTCCTCCCT 115

RESULT 4
LOCUS F14808 126 bp mRNA linear EST 09-SEP-1996
DEFINITION SSC20D02 Porcine small intestine cDNA library Sus scrofa cDNA clone
c20d02, mRNA sequence.
ACCESSION F14808
VERSION F14808.1 GI:971822
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 126)
Wintoro, A.K., Fredholm, M. and Davies, W.
Evaluation and characterization of a porcine small intestine cDNA
library: analysis of 839 clones
Mamm. Genome 7 (7), 509-517 (1996)
Contact: A.K. Wintoro
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowsvej 13, 1870 Frederiksberg C, Denmark.
Location/Qualifiers

source	1. 126	/organism="Sus scrofa"
	/db_xref="taxon:9823"	
	/clone="c20d02"	
	/clone_id="Porcine small intestine CDNA library"	
	/note="directionally cloned cDNA in X1-blue MRF"	
BASE COUNT	9 a 54 c 37 g 24 t	2 others
ORIGIN		
Query Match	90.4%	Score 24.4; DB 10; Length 126;
Best Local Similarity	96.2%;	Pred. No. 3.1e+02;
Matches	25; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Oy	2	cccttcgtctgcctctctccct 27
Db	99	cccttcgtctgcctctctccct 124
RESULT	5	
LOCUS	A0911478	
DEFINITION	LMJFV1_lm85b07.y1 Leishmania major FV1 random genomic library	
ACCESSION	A0911478	
VERSION	A0911478.1	GI:6507994
KEYWORDS	GSS.	
SOURCE	Leishmania major.	
ORGANISM	Leishmania major	
	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;	
REFERENCE	1 (bases 1 to 385)	
AUTHORS	Akopyants, N.S., Clifton, S.W., Martin, J., Page, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Maira, M., Hillier, L., Chinwalla, A., Blistain, A., Schmitt, A., Persson, B., Theising, B., Riteyer, E., Ronko, I., Bennett, J., Cole, R., Anderson, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, R., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S.M.	
TITLE	A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling	
JOURNAL	Mo. Biochem. Parasitol. 113 (2), 337-340 (2001)	
MEDLINE	21192569	
COMMENT	Contact: Akopyants, NS / Beverley, SM	
	Washington University School of Medicine	
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA	
	Tel: 314 286 1800	
	Fax: 314 286 1810	
	Email: est@watson.wustl.edu	
	Library construction: Natalia S. Akopyants, Ph.D.	
	DNA Sequencing by: Washington University Genome Sequencing Center	
	If using this information please cite:	
	N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@porcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@porcim.wustl.edu)	
	Seq primer: -40RP from GIDCO	
	Class: Shotgun	
	High quality sequence status: 383.	
FEATURES	Location/Qualifiers	
source	1. 385	
	/organism="Leishmania major"	
	/strain="Friedlin strain VI"	
	/db_xref="taxon:5664"	
	/clone="LMJFV1_lm85b07"	
	/clone_id="Leishmania major FV1 random genomic library"	
	/lab_host="TOP10 (Invitrogen)"	
	/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;	
	Genomic DNA was isolated from stationary phase cells. For	
	this library, DNA was sheared to give a 10nt size	

BASE COUNT	79 a	141 c	102 g	63 t	
ORIGIN					distribution of 1-1.5kb fragments, blunt-ended with 74 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."
Query Match	Best Local Similarity	84.4%	Score 22.8;	DB 12;	Length 385;
Matches	24: Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
OY	2	cccttcgtcgtcgtccttccttcctccct	27		
Db	319	CCCCGCGCGCGCTCTCTCTCTCTCTCT	344		
RESULT 6					
LOCUS	BE814445	192 bp	mRNA	linear	EST 21-SEP-2000
DEFINITION	MRO-BN0070-290600-020-a05 BN0070	Human sapiens	CDNA,	mRNA sequence.	
ACCESSION	BE814445				
VERSION	BE814445.1	GI:10246679			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 192) Dias Neto,E., Garcia Correa,R., Veijovskl-Almeida,S., Bionesi,M.R., Nagai,M.A., da Silva,W.J.T., Zago,M.A., Bordin,S., Costa,F.F., Godman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunslein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	2020263				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MRO-BN0070-290600-020-a05&t3=2000-06-29&t4=1) Seq primer: puc 18 forward High quality sequence stop: 192. Location/Qualifiers 1..192 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BN0070" /dev_stage="Adult" /note="Organ: breast_normal; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."				
FEATURES					
source	58 a 33 c 54 g 47 t				
BASE COUNT					
ORIGIN					
Query Match	Best Local Similarity	82.2%	Score 22.2;	DB 10;	Length 192;
Matches	24: Conservative	0;	Mismatches	3;	Indels 0; Gaps 0;

OY 1 ccccttgctgcgtcctctccccc 27
||||| ||||||| ||||||| ||||| |||
Db 58 CCCCATCGTCGTCCTCTCTCCT 32

RESULT 7
BF515471/c 472 bp mRNA linear EST 07-DEC-2000
LOCUS IMAGE:3082848 3', mRNA sequence.
DEFINITION UI-H-BMI-ann-e-01-0-UI-s1 NCI-CGAP_Sub7 Homo sapiens cDNA clone
ACCESSION BF515471
VERSION BF515471.1 GI:11600650
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 472)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares lab Clone distribution: NCI-CGAP clone distribution
information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/Image/Image.html
Seq primer: M13 Forward
POLYA=No.

FEATURES
SOURCE

1..472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3082848"
/lab_host="NCI-CGAP_Sub7"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; NCI-CGAP_Sub7
is a subcloned library derived from NCI-CGAP_Sub6. The
NCI-CGAP_Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI-CGAP_Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI-CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1323376-1323911, 1456008-1456773, 1500552-1502855);
NCI-CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI-CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones) 1414920-1417991, 1520904-1522439
); NCI-CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Clones) 1257096-1258631, 1469064-1470983
), 1473592-1476743); NCI-CGAP_Pt22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clones) 985608-986759,
1101192-1101959, 1217928-1220615); NCI-CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clones) 17416-1061255
), 114584-1145351). (6% of the driver population), plus a
pool of 3.840 arrayed clones from NCI-CGAP_Sub1 (IMAGE
Clones) 2708616-2710535) and NCI-CGAP_Sub2 (IMAGE
Clones) 2710536-2712455) (4% of the driver population
, plus a pool of 11,136 clones from NCI-CGAP_Sub3 (IMAGE
Clones) 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI-CGAP_Sub4 (IMAGE
Clones) 2723592-2729326) (40% of the driver population),
plus a pool of 4032 clones from NCI-CGAP_Sub6 (IMAGE
Clones) 2728969-2733190) (40% of the driver population).
Subtraction was performed as previously described [Bonaldo
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_SEQ=None found".

BASE COUNT

110 a 125 c 143 g 94 t

ORIGIN

Query Match 82.2%; Score 22.2; DB 10; Length 472;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ccccttgctgcgtcctctccccc 27
||||| ||||||| ||||||| ||||| |||
Db 311 CCCCATCGTCGTCCTCTCTCCT 285

RESULT 8
A1423306/c 473 bp mRNA linear EST 09-MAR-1999
LOCUS t150905.x1 NCI-CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2102744 3',
DEFINITION mRNA sequence.
ACCESSION A1423306
VERSION A1423306.1 GI:4269237
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 473)
AUTHORS NCI/NIHDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/Image/Image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
Location/Qualifiers

FEATURES
SOURCE

1..473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2102744"
/clone_lib="NCI CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

91 a 142 c 144 g 96 t

Query Match 82.2%; Score 22.2; DB 9; Length 473;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 ccccttgctgcgtcctctccccc 27
||||| ||||||| ||||||| ||||| |||
Db 295 CCCCATCGTCGTCCTCTCTCCT 269

RESULT 9
AM452627 548 bp mRNA linear EST 17-FEB-2000
LOCUS
DEFINITION
UI-H-B13-aju-e-01-0-UI.s1 NC1_CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:3068640 3', mRNA sequence.
AM452627
ACCESSION
AM452627.1 GI:6993403
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 548)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: CGAPBS-Email.nih.gov
Oligo-dt track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/dbp/Image/Image.html
Seq primer: M13 Forward
POLYA-No.

FEATURES
source
1..548
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3068640"
/clone_lib="NCI-CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; NCI-CGAP_Sub5
is a subtracted library derived from NCI-CGAP_Sub4. The
NCI-CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI-CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI-CGAP_Kid3 pool 1 LLAM
3334-3337 3682-3683 3798-3803 (IMAGE Clonids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI-CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clonids 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI-CGAP_Lus pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clonids 1414920-1417991, 1520904-1522439
); NCI-CGAP_GC4 pool 1 LLAM 3164-3167 3716-3720,
3733-3735 (IMAGE Clonids 1257096-1258631, 1469064-1470983,
1473592-1476743); NCI-CGAP_Pt22 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Clonids 985608-986759
1101192-1101959, 1217928-1220615); NCI-CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clonids 1057416-1061255
1144584-1145351). (10% of the driver population), plus a
pool of 3,840 arrayed clones from NCI-CGAP_Sud1 (IMAGE
Clonids 2708616-2710535) and NCI-CGAP_Sub2 (IMAGE
Clonids 2710536-2712455) (10% of the driver population
, plus a pool of 11,136 clones from NCI-CGAP_Sub3 (IMAGE
Clonids 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI-CGAP_Sub4 (IMAGE
Clonids 2723592-2728969) (70% of the driver population).
Subtraction was performed as previously described [Bonaldi
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI-CGAP_Lus
TAG_TISSUE=Lung
TAG_SEQ=CACAC"

BASE COUNT 143 a 140 c 161 g 104 t
ORIGIN

Query Match 82.2%; Score 22.2; DB 9; Length 548;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ccccttcgtcgtccttcctccct 27
|||||
Db 311 CCCCATCGTCGTCTCTCTCTCTCT 285

RESULT 10
BF612947 680 bp mRNA linear EST 14-DEC-2000
LOCUS
DEFINITION
dd79c05.x2 Wellcome CRC pCDNA1 egg Xenopus laevis cDNA clone
IMAGE:3430281 3' similar to SW:FMFL XENLA P5113 FRATILE X MENTAL
RETARDATION PROTEIN 1 HOMOLOG ;, mRNA sequence.
BF612947
ACCESSION
BF612947.1 GI:11784089
KEYWORDS
EST.
SOURCE
African clawed frog.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 680)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
JOURNAL
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
Gurdon. (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: infoimage.llnl.gov
Seq primer: -40up from glbco
High quality sequence stop: 448.

FEATURES
source
1..680
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3430281"
/clone_lib="Wellcome CRC pCDNA1 egg"
/tissue_type="egg"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pCDNA1; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Library was
constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute)"

BASE COUNT 146 a 162 c 128 g 243 t
ORIGIN

Query Match 82.2%; Score 22.2; DB 10; Length 680;
Best Local Similarity 88.9%; Pred. No. 1.7e+03;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ccccttcgtcgtccttcctccct 27
|||||
Db 573 CCCCTCGTCGTCTCTCTCTCT 599

RESULT 11
CNS0164L 1201 bp DNA linear GSS 26-JUL-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN15C21 of DrosBAC library from Drosophila melanogaster (fruit

ACCESSION	fly), genome survey sequence.
VERSION	AL106287
KEYWORDS	AL106287.1 GI:5621177
SOURCE	GSS.
ORGANISM	fruit fly.
REFERENCE	Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
TITLE	1 (bases 1 to 1201)
JOURNAL	Genoscope.
COMMENT	Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (PDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.
FEATURES	location/Qualifiers
source	1..1201
	/organism="Drosophila melanogaster" /plasmid="pBelOBAC11" /db_xref="taxon:7227" /clone_11b="DrosBAC" /clone="BACN15C21" /note="end : SP6"
BASE COUNT	304 a 204 c 230 g 275 t 188 others
ORIGIN	
Query Match	82.2%; Score 22.2; DB 12; Length 1201;
Best Local Similarity	77.8%; Pred. No. 1.8e+03;
Matches 21; Conservative 4; Mismatches 2; Indels 0; Gaps 0;	
Oy 1	ccccctgctgctctctccccc 27
Db 1032	:: :: ::
CCCCCTTSTCTCTCTCTCTCTCCCT 1006	
RESULT 12	
BM324144	
LOCUS	BM324144 513 bp mRNA linear EST 04-JAN-2002
DEFINITION	PtC1_24.F12.b1.A002 pathogen-infected compatible 1 (PtC1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION	BM324144
VERSION	BM324144.1 GI:18062572
KEYWORDS	EST.
SOURCE	sorghum.
ORGANISM	Sorghum bicolor
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 513)
AUTHORS	Cordonnier-Pratt,M.-M., Glinge,A., Fang,G.C., Dean,R., Wing,R., Sudman,M. and Pratt,L.H.
TITLE	An EST database from Sorghum: plants infected with a compatible pathogen
JOURNAL	Unpublished (2002)
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel.: 706 542 1800 Fax: 706 542 1805 Email: mmp1att@uga.edu Sequences have been trimmed to exclude PolyA, vector, and regions below phred quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with PolyTmix or

```

FEATURES
  source
    1. 513
      Location/Qualifiers
        /organism="Sorghum bicolor"
        /cultivar="BRX623"
        /db_xref="taxon:4558"
        /clone_lib="Pathogen-infected compatible 1 (PIC1)"
        /tissue_type="Leaves"
        /dev_stage="4-week-old seedlings infected with
        Colletotrichum graminicola"
        /note="Vector: pBluescript, II SK(-) from lambda Zap II;
        site 1: XhoI; site 2: EcoRI. Four-week-old sorghum
        seedlings were sprayed with spore suspension prepared from
        3-week-old FRM421, a sorghum isolate of the anthracose
        pathogen Colletotrichum graminicola. Inoculated plants
        were kept in a 25 C dark growth chamber with 100% relative
        humidity for 24 hr, followed by 12/12 hr of light/dark
        cycle at 25 C with 90% relative humidity for another 24
        hr. All leaves were harvested and quick frozen with liquid
        nitrogen and stored in a -80 C freezer. The library was
        made from poly-A RNA in the cloning vector lambda Zap II.
        Clones to be sequenced were prepared by mass excision.
        WARNING: While most or all ESTs are expected to derive
        from the host plant, no effort was made to eliminate ESTs
        deriving from the pathogen."

BASE COUNT
  126 a 143 c 144 g 100 t

ORIGIN
  1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Query Match 79.3%, Score 21.4; DB 10; Length 513;
Best Local Similarity 95.7%; Pred. No. 2.9e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Db 79 cccctcgctgcgtccctcc 101

RESULT 13
BF277889/c 538 bp mRNA linear EST 07-MAR-2001
LOCUS BF277889/c
DEFINITION GA_ED031P15f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboresum cDNA clone GA_ED031P15f, mRNA sequence.
ACCESSION BF277889
VERSION BF277889.1 GI:11208959
KEYWORDS EST.
ORGANISM Gossypium arboreum.
SOURCE Gossypium arboreum.
  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 538)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: twing@clemson.edu
  Seq primer: TAATACGACTCACAATAGCG
  High quality sequence start: 31
  High quality sequence stop: 504.
  Location/Qualifiers
    1. 538
      /organism="Gossypium arboreum"
FEATURES
  source
    1. 538
      Location/Qualifiers
        /organism="Gossypium arboreum"
        /cultivar="BRX623"
        /db_xref="taxon:4558"
        /clone_lib="Pathogen-infected compatible 1 (PIC1)"
        /tissue_type="Leaves"
        /dev_stage="4-week-old seedlings infected with
        Colletotrichum graminicola"
        /note="Vector: pBluescript, II SK(-) from lambda Zap II;
        site 1: XhoI; site 2: EcoRI. Four-week-old sorghum
        seedlings were sprayed with spore suspension prepared from
        3-week-old FRM421, a sorghum isolate of the anthracose
        pathogen Colletotrichum graminicola. Inoculated plants
        were kept in a 25 C dark growth chamber with 100% relative
        humidity for 24 hr, followed by 12/12 hr of light/dark
        cycle at 25 C with 90% relative humidity for another 24
        hr. All leaves were harvested and quick frozen with liquid
        nitrogen and stored in a -80 C freezer. The library was
        made from poly-A RNA in the cloning vector lambda Zap II.
        Clones to be sequenced were prepared by mass excision.
        WARNING: While most or all ESTs are expected to derive
        from the host plant, no effort was made to eliminate ESTs
        deriving from the pathogen."

BASE COUNT
  126 a 143 c 144 g 100 t

ORIGIN
  1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Query Match 79.3%, Score 21.4; DB 10; Length 513;
Best Local Similarity 95.7%; Pred. No. 2.9e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Db 79 cccctcgctgcgtccctcc 101

RESULT 13
BF277889 538 bp mRNA linear EST 07-MAR-2001
LOCUS BF277889
DEFINITION GA_ED031P15f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboresum cDNA clone GA_ED031P15f, mRNA sequence.
ACCESSION BF277889
VERSION BF277889.1 GI:11208959
KEYWORDS EST.
ORGANISM Gossypium arboreum.
SOURCE Gossypium arboreum.
  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 538)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: twing@clemson.edu
  Seq primer: TAATACGACTCACAATAGCG
  High quality sequence start: 31
  High quality sequence stop: 504.
  Location/Qualifiers
    1. 538
      /organism="Gossypium arboreum"
FEATURES
  source
    1. 538
      Location/Qualifiers
        /organism="Gossypium arboreum"
        /cultivar="BRX623"
        /db_xref="taxon:4558"
        /clone_lib="Pathogen-infected compatible 1 (PIC1)"
        /tissue_type="Leaves"
        /dev_stage="4-week-old seedlings infected with
        Colletotrichum graminicola"
        /note="Vector: pBluescript, II SK(-) from lambda Zap II;
        site 1: XhoI; site 2: EcoRI. Four-week-old sorghum
        seedlings were sprayed with spore suspension prepared from
        3-week-old FRM421, a sorghum isolate of the anthracose
        pathogen Colletotrichum graminicola. Inoculated plants
        were kept in a 25 C dark growth chamber with 100% relative
        humidity for 24 hr, followed by 12/12 hr of light/dark
        cycle at 25 C with 90% relative humidity for another 24
        hr. All leaves were harvested and quick frozen with liquid
        nitrogen and stored in a -80 C freezer. The library was
        made from poly-A RNA in the cloning vector lambda Zap II.
        Clones to be sequenced were prepared by mass excision.
        WARNING: While most or all ESTs are expected to derive
        from the host plant, no effort was made to eliminate ESTs
        deriving from the pathogen."

BASE COUNT
  126 a 143 c 144 g 100 t

ORIGIN
  1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Query Match 79.3%, Score 21.4; DB 10; Length 513;
Best Local Similarity 95.7%; Pred. No. 2.9e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Db 79 cccctcgctgcgtccctcc 101

RESULT 13
BF277889 538 bp mRNA linear EST 07-MAR-2001
LOCUS BF277889
DEFINITION GA_ED031P15f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboresum cDNA clone GA_ED031P15f, mRNA sequence.
ACCESSION BF277889
VERSION BF277889.1 GI:11208959
KEYWORDS EST.
ORGANISM Gossypium arboreum.
SOURCE Gossypium arboreum.
  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 538)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: twing@clemson.edu
  Seq primer: TAATACGACTCACAATAGCG
  High quality sequence start: 31
  High quality sequence stop: 504.
  Location/Qualifiers
    1. 538
      /organism="Gossypium arboreum"
FEATURES
  source
    1. 538
      Location/Qualifiers
        /organism="Gossypium arboreum"
        /cultivar="BRX623"
        /db_xref="taxon:4558"
        /clone_lib="Pathogen-infected compatible 1 (PIC1)"
        /tissue_type="Leaves"
        /dev_stage="4-week-old seedlings infected with
        Colletotrichum graminicola"
        /note="Vector: pBluescript, II SK(-) from lambda Zap II;
        site 1: XhoI; site 2: EcoRI. Four-week-old sorghum
        seedlings were sprayed with spore suspension prepared from
        3-week-old FRM421, a sorghum isolate of the anthracose
        pathogen Colletotrichum graminicola. Inoculated plants
        were kept in a 25 C dark growth chamber with 100% relative
        humidity for 24 hr, followed by 12/12 hr of light/dark
        cycle at 25 C with 90% relative humidity for another 24
        hr. All leaves were harvested and quick frozen with liquid
        nitrogen and stored in a -80 C freezer. The library was
        made from poly-A RNA in the cloning vector lambda Zap II.
        Clones to be sequenced were prepared by mass excision.
        WARNING: While most or all ESTs are expected to derive
        from the host plant, no effort was made to eliminate ESTs
        deriving from the pathogen."

BASE COUNT
  126 a 143 c 144 g 100 t

ORIGIN
  1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Query Match 79.3%, Score 21.4; DB 10; Length 513;
Best Local Similarity 95.7%; Pred. No. 2.9e+03;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cccctcgctgcgtccctcc 23
  |||||||||||||||
  79 cccctcgctgcgtccctcc 101

Db 79 cccctcgctgcgtccctcc 101

RESULT 13
BF277889 538 bp mRNA linear EST 07-MAR-2001
LOCUS BF277889
DEFINITION GA_ED031P15f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboresum cDNA clone GA_ED031P15f, mRNA sequence.
ACCESSION BF277889
VERSION BF277889.1 GI:11208959
KEYWORDS EST.
ORGANISM Gossypium arboreum.
SOURCE Gossypium arboreum.
  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eustosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 538)
  Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
  D., Wood, T.C., Leslie, A. and Wilkins, T.A.
  An integrated analysis of the genetics, development, and evolution
  of the cotton fiber
  Unpublished (2000)
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 728
```

```

/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_EB003P15f"
/clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/Note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      95 a      190 c      196 g      53 t      4 others
ORIGIN

Query Match      78.5%; Score 21.2; DB 10; Length 538;
Best Local Similarity 88.5%; Pred. No. 3.3e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccctgcgtcgtcctcctcctcctcct 27
    ||| ||||| ||||| ||||| |||

DB 240 CCCCTGCTGCTGCTCCTCCTCCT 215

RESULT 14      575 bp      mRNA      linear      EST 22-OCT-2001
LOCUS      BF616867/c
DEFINITION      HVSMC0013E04f Hordeum vulgare seedling shoot EST library
                  HVCDNA0003 (Etolated and unstressed) Hordeum vulgare cDNA clone
ACCESSION      BF616867
VERSION      BF616867
KEYWORDS      EST.
SOURCE      BF616867.2 GI:13108437
ORGANISM      Hordeum vulgare
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
                ; Triticeae; Hordeum.
REFERENCE      1 (bases 1 to 575)
AUTHORS      Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
                ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
                ,R.D., Oates,R. and Main,D.
                Development of a genetically and physically anchored EST resource
                for barley genomics: Morex unstressed seedling shoot cDNA library
                Unpublished (2001)
                On Dec 18, 2000 this sequence version replaced gi:11880601.
TITLE      Contact: Wing RA
                Clemson University Genomics Institute
                Clemson University
                100 Jordan Hall, Clemson, SC 29634, USA
                Tel: 864 656 7288
                Fax: 864 656 4293
                Email: rwing@clemson.edu
                Total hg bases = 194
                Seq primer: AATTACCTCCTCACTAAAGG
                High quality sequence stop: 565.
                Location/Qualifiers
                1..575
FEATURES
SOURCE
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMC0013E04f"
/clone_1lb="Hordeum vulgare seedling shoot EST library
HVCDNA0003 (Etolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/Note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling shoots were
then harvested, total RNA was prepared; poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give phuscrypt
SK(-) cDNA phagemids. These steps were performed in the T3

```

```

Close laboratory at the University of California,
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9paps/bgn/31/clover.html)"

BASE COUNT      106 a      193 c      196 g      77 t      3 others
ORIGIN

Query Match      78.5%; Score 21.2; DB 10; Length 575;
Best Local Similarity 88.5%; Pred. No. 3.4e+03;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cccctgcgtcgtcctcctcctcctcct 27
    ||| ||||| ||||| ||||| |||

DB 168 CCCCTGCTGCTGCTCCTCCTCCT 143

RESULT 15      629 bp      mRNA      linear      EST 22-OCT-2001
LOCUS      BG300411/c
DEFINITION      HVSMB0016N18f Hordeum vulgare seedling shoot EST library
                  HVCDNA0002 (dehydration stress) Hordeum vulgare cDNA clone
ACCESSION      BG300411
VERSION      BG300411
KEYWORDS      EST.
SOURCE      BG300411.1 GI:13097938
ORGANISM      Hordeum vulgare
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
                ; Triticeae; Hordeum.
REFERENCE      1 (bases 1 to 629)
AUTHORS      Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
                ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
                ,R.D., Oates,R. and Main,D.
                Development of a genetically and physically anchored EST resource
                for barley genomics: Morex drought-stressed seedling shoot cDNA
                library
                Unpublished (2001)
                Contact: Wing RA
                Clemson University Genomics Institute
                Clemson University
                100 Jordan Hall, Clemson, SC 29634, USA
                Tel: 864 656 7288
                Fax: 864 656 4293
                Email: rwing@clemson.edu
                Total hg bases = 230
                Seq primer: AATTACCTCCTCACTAAAGG
                High quality sequence stop: 610.
                Location/Qualifiers
                1..629
FEATURES
SOURCE
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMB0016N18f"
/clone_1lb="Hordeum vulgare seedling shoot EST library
HVCDNA0002 (dehydration stress)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/Note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic

```


conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were incubated at 90% RH for 24 hr. Shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, 600000 pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)

BASE COUNT 125 a 211 c 196 g 97 t

ORIGIN

Query Match

Best Local Similarity 88.5%; Score 21.2; DB 10; Length 629;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cccttcgctgcgtcctctccct 27
||| ||||| ||||| ||||| |||||
DB 193 CCCCTGCTGCTGCTGCTCTCTCT 168

Search completed: August 14, 2002, 21:04:48
Job time: 11036 sec